抽样论

许宝騄 著

北京大学数学丛书

抽 样 论

许 宝 騄 著

北 京 大 学 出 版 社

内 容 简 介

本书所述之抽样论，即所谓大规模调查抽样论，有很大的实用价值。本书篇幅很少，但却包括了各种抽样法的主要理论。叙述简明清晰，定理的证明浅显易懂。具有微积分和初等概率统计知识的读者均可读懂。

全书共分五章，包括：概论、随机抽样法、分层抽样法、二阶抽样法、以及集团抽样法和系统抽样法。

本书可供高等院校数学系高年级学生及研究生选修课的教材，亦可供从事抽样调查的实际工作者及数理统计工作者参考。

抽 样 论

北京大学出版社出版
（北京大學校内）
新华书店北京发行所发行
国防科委印刷厂印刷
850×1168毫米 32开本 2.625印张 59千字
1982年4月第一版 1982年4月第一次印刷
印数1—32,000册

统一书号：13209·37 定价：0.42元

出 版 说 明

本书所述之抽样论，即所谓大规模调查抽样论，有很大的实用价值。常应用于人口调查，能源调查，社会经济调查，森林、草原、农田估产，昆虫计数等方面。在四个现代化建设中，我们的各项工作都应符合客观规律，不能主观盲目。这首先就要了解客观情况，进行调查研究。调查中如能正确使用抽样调查，将会收到事半功倍、多快好省的效果。

一九六〇年前后，北京大学数学系概率统计专门化的同学，曾从事这方面的实际工作。为此许宝騄教授写了"抽样论讲义"。该讲义的取材主要参考了 W·G·Cochran 著 "*Sampling Techniques*" 一书。许先生的讲义篇幅虽小，但却包括了各种抽样法的主要理论。内容叙述简明清晰，有关定理的证明均浅显易懂。具有微积分和初等概率统计知识的读者即可读懂。

本书前四章介绍了随机抽样、分层抽样、二阶抽样等一些基本的抽样方法，并介绍了相应的估值法的理论依据。这四章是根据许宝騄先生的讲义稍加整理而成。许先生在概论中提到集团抽样与系统抽样，但讲义后面的章节中没有这方面的内容。虽然这两种抽样法从数学理论上可视为二阶抽样的特例，但这两种抽样法简便易行，常为实际工作者采用。为便于实际工作者参考，由孙山泽同志参照前四章的体例，增写了第五章集团抽样法和系统抽样法。这一章的取材也参考了 "*Sampling Techniques*" 一书。

北京大学数学系概率统计教研室

一九八一年三月

说　　明

此丛书是以数学、计算数学、概率统计及有关专业的高年级、研究生、青年教师及数学研究工作者为读者对象的出版物。丛书特点是内容新颖，力图反映现代数学的新成就；叙述精练，约相当于一学期三学时研究生课程的取材。我们编辑出版此丛书的主要目的是为了适应我们国家培养研究生的需要，同时，又可作为数学及有关系科高年级选修课程的参考书，为提高本科生的教学质量贡献一份力量。

我们诚恳地希望：广大读者对于书目的选择，内容的取材提出宝贵意见，作为我们今后出版或再版时的参考。

《北京大学数学丛书》编委会

一九八一年元月

目　　录

第一章 概　　论

抽样论亦称大规模调查抽样论，常用于人口调查、能源调查、社会经济调查、森林林木估积、草原及农田估产、昆虫数量估计等方面.

抽样调查有极大的实用价值. 在正确理论指导下，合理地抽样既可获得可靠的高精度的信息，又可大大节约调查的人力、物力、财力、时间. 特别地，当信息有很强的时间性时，旷日持久的全面调查将只能获得陈旧的信息，而毫无价值，因而必须进行抽样调查.

§1.1 问题的基本提法

设 π_N 是由 N 个有明确编号的个体 O_1，O_2，…，O_N 组成之有限总体（O_1，O_2，…，O_N 是抽象的个体，今后不妨以其号码 1，2，…，N 表之）. 在 π_N 中每个个体对应一数量指标，依次为 Y_1，Y_2，…，Y_N，其中 N 是已知正整数. 用各种抽样办法从 π_N 中抽取 n 个个体，测其数量指标，得 y_1，y_2，…，y_n（y_1，y_2，…，y_n 系 Y_1，Y_2，…，Y_N 的一部分）. 据此估计总体 π_N 的下列数字特征①：

(i) $\bar{Y}=\frac{1}{N}\sum_{i=1}^{N} Y_i$ （或 $\tilde{Y}=\sum_{i=1}^{N} Y_i$）；

---

① 设想随机变量 Y，其可能值为 $\{Y_1, Y_2, \cdots, Y_N\}$.

$$P\{Y=Y_i\}=\frac{1}{N},\quad i=1,\ 2,\ \cdots,\ N$$

则 $EY=\bar{Y}$，$\mathrm{Var}(Y)=\sigma^2$，故称 $\bar{Y}$ 为 π_N 的期望，σ^2 为方差.

(ii) $\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(Y_i-\bar{Y})^2$ $\left(或\ S^2=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-\bar{Y})^2\right)$;

(iii) $C^2=S^2/\bar{Y}^2$ (变异系数).

并研究估计的误差及最小抽样数目等问题.

例 一块100平方公里的地方，是蝗虫的幼虫蝗蝻的出生地，今要调查蝗蝻本年的出生量以便开展治蝗工作，为此将这块地方设想分成100000000小块，每小块一平方米. 调查其中的1000个小块，计算这些小块中蝗蝻数. 得1000个数据 $y_1, y_2, \cdots, y_{1000}$. 这里 $N=100000000$，该地块的蝗蝻总量为 $\tilde{Y}=Y_1+Y_2+\cdots+Y_{100000000}$，抽样数为 $n=1000$，测得的1000个样的指标即为 $y_1, y_2, \cdots, y_{1000}$，要据此估计 $\tilde{Y}$.

§1.2 各种抽样法列举

（一）随机抽样法（亦称无限制抽样法）

在 π_N 的 N 个个体中机会均等地抽取第一个样，然后在剩下的$(N-1)$个个体中机会均等地抽取第二个样，……，最后，在所剩 $N-(n-1)$ 个个体中机会均等地抽取第 n 个样（即所谓等概无放回抽样），测每个样的指标. 我们称这种抽样法为随机抽样法. 被抽中之个体称为入样个体. 具体工作，可将 N 个个体与号码1，2，…，N 建立对应，通过抽取对应的号码来实现.

（二）分层抽样法

将 π_N 先分成 K 组：$\pi_{N_1}, \pi_{N_2}, \cdots, \pi_{N_K}$. 称 π_{N_i} 为第 i 层. 把每一层看作一个小总体，对之抽取一组随机样本. 这 K 组样本合成 π_N 之分层样本.

当 π_N 中某些个体有明显差异时，则应将相近的个体归为一组. 如此将 π_N 分成 K 组，采用分层抽样法取样.

（三）二阶抽样法

将 π_N 分成 K 组：$\pi_{N_1}, \pi_{N_2}, \cdots \pi_{N_K}$. 这些组称为 π_N 的第一性抽样单位，将它们看作 K 个个体进行抽样. 然后再对抽得之

组 π_{N_i} 内之个体抽样．π_{N_i} 中的个体称为 π_{N} 的第二性抽样单位．

当 π_N 中个体因数量太大，或其它技术上的原因，无法直接编号时，则可采用二阶抽样法，先按组编号，抽取若干组，再在抽得的组内将个体编号，抽取个体．

（四）多阶抽样法

设想(三)中第二性抽样单位仍不是 π_N 的个体，而是一小组，对它再作第三性抽样．余此类推．

（五）集团抽样法

将 π_N 分成 K 组：π_{N_1}，π_{N_2}，…，π_{N_K}．对这 K 个组进行抽样，对入样的 π_{N_i} 不再对其中的个体进行抽样，而是将其中的个体逐个观测．第 i 组 π_{N_i} 称为第 i 个集团．

（六）系统抽样法（又称机械抽样法）

选一正整数 K，将 π_N 中个体逐个排列如下：

$$
\begin{array}{cccc}
1, & 2, & \cdots, & K, \\
K+1, & K+2, & \cdots, & 2K, \\
2K+1, & 2K+2, & \cdots, & 3K, \\
\cdots\cdots & & & \\
\end{array}
$$

直至排到 N 为止

对号码 1，2，…，K 作随机抽样（常常只抽一个），若 i 入样，则 $K+i$，$2K+i$，$3K+i$，…皆入样．

以下各章将分别讨论各种抽样法．

第二章　随机抽样法

这一方法是等概地从总体 π_N 中无放回抽取 n 个样．具体的实现，可将 π_N 的 N 个个体标以号码 1，2，…，N．然后利用随机数表抽出 n 个不同的数码，或用计算机产生随机数来抽取 n 个不同数码．用随机数表抽取的方法如下：设 π_N 是 $N=345$，要抽 $n=15$ 个样．则从随机数表中任取的三列所构成的三位数中，依次取出不同的三位数，当数在 001—345 之间时，则该号码入样，当数在401—745之间时，则该数减去 400 的号码入样，其余的 000，346—400，746—999 不要．当某号码已入样，而再次碰到该号码入样时，则只算一次．例如，用下列随机数表取样：

随　机　数　表

65	54	73	88	44	76	68	47	93	11	14	95	72	94	14
95	84	67	58	37	62	18	03	23	61	60	88	44	62	99
05	63	05	42	44	63	44	78	98	09	25	58	00	57	12
36	05	60	21	12	26	61	99	62	44	83	09	78	74	84
53	45	44	36	44	78	74	09	25	58	00	57	12	38	46
05	60	21	13	26	61	99	62	44	78	73	55	63	86	00
58	22	57	86	27	63	43	45	60	74	16	72	73	79	50
79	59	69	50	72	47	26	95	60	88	44	83	09	79	49
69	39	80	45	69	92	93	38	12	57	00	58	16	23	54
56	32	36	22	58	69	50	72	47	26	61	99	87	90	83

77 93 81 16 61	05 97 78 84 43	59 76 16 76 91
29 63 99 16 65	60 82 94 19 25	98 28 24 57 05
61 10 39 10 58	48 81 77 60 31	45 17 09 62 16
34 31 36 56 98	31 26 20 03 27	63 09 56 36 16
35 16 11 17 56	31 58 25 87 90	05 29 87 77 34

则可先利用 8，9，10 列构成三位数．第一个数为844，舍去．第二个数为 837．也舍去．第三个数为 244，则244号入样．第四个为112，112号入样．第五个为644，减去400为244，由于244号已入样，故该数也舍去．如此继续，可得 326 号、227 号、72 号、169号、258号、261号、265号、58号、298号入样．8，9，10列随机数已取完，但入样号码仍不足15个，此时可转到另三列继续取．如利用18，19，20列，则继续得311号、158号、76号、74号入样．取满15个即停止．

实际工作中，有时从一个装有号码 1—N 的口袋中，摸出号码来实现抽样．

§2.1 预备定理

定义2.1 设$[Y_1, Y_2, \cdots, Y_N]$为总体之数量指标，经随机抽样测得 $y_1, y_2, \cdots, y_n$，则称 $(y_1, y_2, \cdots, y_n)$ 是来自总体$[Y_1, Y_2, \cdots, Y_N]$的随机样本．显然，任一 y_i 必等于总体中某一Y_{θ_i}．θ_i 称为第 i 个入样号码．

定理2.1 令 θ_i 为随机抽样的第 i 个入样号码，$i=1,2\cdots, n$，则对任一组指定的号码 $a_1, a_2, \cdots, a_n$ $(1\leqslant a_i\leqslant N$，且 $a_i\neq a_j$，当 $i\neq j)$，总有

$$P\{\theta_1=a_1,\theta_2=a_2,\cdots,\theta_n=a_n\}$$

$$=\frac{1}{N}\cdot\frac{1}{N-1}\cdot\cdots\cdot\frac{1}{N-n+1}=\frac{(N-n)!}{N!}$$

证明

$$P\{\theta_1=a_1,\ \theta_2=a_2,\ \cdots,\ \theta_n=a_n\}$$
$$=P\{\theta_1=a_1\}\cdot P\{\theta_2=a_2|\theta_1=a_1\}\cdot\cdots\cdot P\{\theta_n=a_n|\theta_1=a_1,\cdots,\theta_{n-1}=a_{n-1}\}$$
$$=\frac{1}{N}\cdot\frac{1}{N-1}\cdot\cdots\cdot\frac{1}{N-n+1}$$

定理2.2 令 $\theta_i(i=1,2,\cdots,n)$ 为随机抽样的第 i 个入样号码．对任意 $m(1\leqslant m\leqslant n)$ 及任意指定的号码 $k_1,\ k_2,\ \cdots,\ k_m(1\leqslant k_j\leqslant n,\ j=1,\ \cdots,\ m;\ k_i\neq k_j$，当 $i\neq j$) 及 $\beta_1,\ \beta_2,\ \cdots,\ \beta_m(1\leqslant\beta_j\leqslant N,\ j=1,\ \cdots,\ m;\ \beta_i\neq\beta_j$，当 $i\neq j$)，总有

$$P\{\theta_{k_1}=\beta_1,\ \theta_{k_2}=\beta_2,\ \cdots,\ \theta_{k_m}=\beta_m\}$$
$$=\frac{1}{N}\cdot\frac{1}{N-1}\cdot\cdots\cdot\frac{1}{N-m+1}=\frac{(N-m)!}{N!}$$

证明 号码 1, 2, …, n 中，除去 $k_1,\ k_2,\ \cdots,\ k_m$，还余 $n-m$ 个，记作 $l_1,\ l_2,\ \cdots,\ l_{n-m}$．于是

$$P\{\theta_{k_1}=\beta_1,\ \cdots,\ \theta_{k_m}=\beta_m\}$$
$$=\sum_G P\{\theta_{k_1}=\beta_1,\ \cdots,\ \theta_{k_m}=\beta_m,\ \theta_{l_1}=\gamma_1,\ \cdots,\ \theta_{l_{n-m}}=\gamma_{n-m}\}$$

这里 $\sum_G$ 表示对 $\gamma_1,\ \cdots,\ \gamma_{n-m}$ 求和，且 $1\leqslant\gamma_j\leqslant N,\ j=1,\ \cdots,\ n-m;\ \gamma_i\neq\gamma_j$，当 $i\neq j$；$\gamma_j\neq\beta_1,\ \cdots,\ \beta_m$．由定理 2.1 知 $\sum_G$ 内每一项皆等于$(N-n)!/N!$，故

$$P\{\theta_{k_1}=\beta_1,\ \cdots,\ \theta_{k_m}=\beta_m\}$$
$$=\sum_G\frac{(N-n)!}{N!}=\frac{(N-n)!}{N!}P_{N-m}^{n-m}$$
$$=\frac{(N-n)!}{N!}\ \frac{(N-m)!}{[(N-m)-(n-m)]!}$$

$$=\frac{(N-m)!}{N!}$$

其中 P_{N-m}^{n-m} 表示由 $N-m$ 元中取 $n-m$ 元之排列数.

系 令 $m=1, 2, 3, 4$, 得

$$P\{\theta_i=\beta\}=\frac{1}{N} \quad (i=1, \cdots, n; \beta=1, \cdots, N)$$

$$P\{\theta_i=\beta_1, \theta_j=\beta_2\}=\frac{1}{N(N-1)} \quad (i\neq j; \beta_1\neq\beta_2)$$

$$P\{\theta_i=\beta_1, \theta_j=\beta_2, \theta_k=\beta_3\}=\frac{1}{N(N-1)(N-2)}$$

(i, j, k 互不相等; β_1, β_2, β_3 互不相等)

$$P\{\theta_i=\beta_1, \theta_j=\beta_2, \theta_k=\beta_3, \theta_l=\beta_4\}$$

$$=\frac{1}{N(N-1)(N-2)(N-3)}$$

(i,j,k,l 互不相等; $\beta_1,\beta_2,\beta_3,\beta_4$ 互不相等)

定理2.3 设 $u_1, \cdots, u_n$ 是来自 $(U_1, \cdots, U_N)$ 的随机样本, 且

$$\bar{U}=\frac{1}{N}\sum_{i=1}^{N} U_i=0$$

则

(i) $$E(\bar{u}^2)=\frac{N-n}{nN(N-1)}\sum_{\beta=1}^{N} U_\beta^2=\left(\frac{1}{n}-\frac{1}{N}\right)S_U^2;$$

(ii) $$E(\bar{u}^3)=\frac{(N-n)(N-2n)}{n^2N(N-1)(N-2)}\sum_{\beta=1}^{N}U_\beta^3;$$

(iii) $$E(\bar{u}^4)=\frac{(N-n)[N^2-(6n-1)N+6n^2]}{n^3N(N-1)(N-2)(N-3)}\sum_{\beta=1}^{N}U_\beta^4$$

$$+\frac{3(n-1)(N-n)(N-n-1)}{n^3N(N-1)(N-2)(N-3)}\left(\sum_{\beta=1}^{N}U_\beta^2\right)^2.$$

证明 证(i)：

$$\bar{u}^2=\left(\frac{1}{n}\sum_{i=1}^{n}u_i\right)^2=\frac{1}{n^2}\sum_{i=1}^{n}u_i^2+\frac{1}{n^2}\sum_{i\neq j}u_iu_j$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}U_{\theta_i}^2+\frac{1}{n^2}\sum_{i\neq j}U_{\theta_i}U_{\theta_j}$$

故

$$E(\bar{u}^2)=\frac{1}{n^2}\sum_{i=1}^{n}E(U_{\theta_i}^2)+\frac{1}{n^2}\sum_{i\neq j}E(U_{\theta_i}U_{\theta_j})$$

由定理 2.2 之系，得

$$E(U_{\theta_i}^2)=\sum_{\beta=1}^{N}U_\beta^2P\{\theta_i=\beta\}=\frac{1}{N}\sum_{\beta=1}^{N}U_\beta^2$$

$$E(U_{\theta_i}U_{\theta_j})=\sum_{\beta\neq\gamma}U_\beta U_\gamma P\{\theta_i=\beta,\ \theta_j=\gamma\}$$

$$=\frac{1}{N(N-1)}\sum_{\beta\neq\gamma}U_\beta U_\gamma,\quad i\neq j$$

又由 $\bar{U}=0$ 得

$$0=(N\bar{U})^2=\left(\sum_{\beta=1}^{N}U_\beta\right)^2=\sum_{\gamma=1}^{N}\sum_{\beta=1}^{N}U_\beta U_\gamma$$

$$=\sum_{\beta=1}^{N}U_\beta^2+\sum_{\beta\neq\gamma}U_\beta U_\gamma$$

故有

$$\sum_{\beta\neq\gamma}U_\beta U_\gamma=-\sum_{\beta=1}^{N}U_\beta^2$$

$$E(U_{\theta_i}U_{\theta_j})=\frac{-1}{N(N-1)}\sum_{\beta=1}^{N}U_\beta^2$$

从而

$$E(\bar{u}^2)=\frac{1}{n^2}\frac{n}{N}\sum_{\beta=1}^{N}U_\beta^2-\frac{1}{n^2}\frac{n(n-1)}{N(N-1)}\sum_{\beta=1}^{N}U_\beta^2$$

$$=\frac{N-n}{nN(N-1)}\sum_{\beta=1}^{N}U_\beta^2$$

证(ii):

$$\bar{u}^3=\frac{1}{n^3}\sum_{i,j,k}u_iu_ju_k$$

$$=\frac{1}{n^3}\sum_{i=1}^{n}u_i^3+\frac{3}{n^3}\sum_{i\neq j}u_i^2u_j+\frac{1}{n^3}\sum_{i\neq j\neq k}u_iu_ju_k$$

由定理2.2之系及$\bar{U}=0$，易证

$$E(u_i^3)=\frac{1}{N}\sum_{\beta=1}^{N}U_\beta^3$$

对于 $i\neq j$

$$E(u_i^2u_j)=\frac{1}{N(N-1)}\sum_{\beta\neq\gamma}U_\beta^2U_\gamma$$

而

$$0=\left(\sum_{\gamma=1}^{N}U_\gamma\right)\left(\sum_{\beta=1}^{N}U_\beta^2\right)=\sum_{\beta=1}^{N}U_\beta^3+\sum_{\beta\neq\gamma}U_\beta^2U_\gamma$$

所以

$$E(u_i^2u_j)=\frac{-1}{N(N-1)}\sum_{\beta=1}^{N}U_\beta^3$$

对于 $i\neq j\neq k$

$$E(u_iu_ju_k)=\frac{1}{N(N-1)(N-2)}\sum_{\beta\neq\gamma\neq\delta}U_\beta U_\gamma U_\delta$$

而

$$0=\left(\sum_{\beta=1}^{N}U_\beta\right)^3=\sum_{\beta=1}^{N}U_\beta^3+3\sum_{\beta\neq\gamma}U_\beta^2U_\gamma+\sum_{\beta\neq\gamma\neq\delta}U_\beta U_\gamma U_\delta$$

$$=\sum_{\beta=1}^{N}U_\beta^3-3\sum_{\beta=1}^{N}U_\beta^3+\sum_{\beta\neq\gamma\neq\delta}U_\beta U_\gamma U_\delta$$

$$=-2\sum_{\beta=1}^{N}U_\beta^3+\sum_{\beta\neq\gamma\neq\delta}U_\beta U_\gamma U_\delta$$

所以

$$E(u_iu_ju_k)=\frac{2}{N(N-1)(N-2)}\sum_{\beta=1}^{N}U_\beta^3$$

从而

$$E(\bar{u}^3)=\frac{1}{n^3}\frac{n}{N}\sum_{\beta=1}^{N}U_\beta^3-\frac{3}{n^3}\frac{n(n-1)}{N(N-1)}\sum_{\beta=1}^{N}U_\beta^3$$

$$+\frac{2}{n^3}\frac{n(n-1)(n-2)}{N(N-1)(N-2)}\sum_{\beta=1}^{N}U_\beta^3$$

$$=\frac{(N-n)(N-2n)}{n^2N(N-1)(N-2)}\sum_{\beta=1}^{N}U_\beta^3$$

证(iii)：

$$\bar{u}^4=\frac{1}{n^4}\sum_{i,j,k,l}u_iu_ju_ku_l$$

$$=\frac{1}{n^4}\sum_{i=1}^{n}u_i^4+\frac{4}{n^4}\sum_{i\neq j}u_i^3u_j+\frac{3}{n^4}\sum_{i\neq j}u_i^2u_j^2$$

$$+\frac{6}{n^4}\sum_{i\neq j\neq k}u_i^2u_ju_k+\frac{1}{n^4}\sum_{i\neq j\neq k\neq l}u_iu_ju_ku_l$$

由定理2.2之系及$\bar{U}=0$，易证

$$E(u_i^4)=\frac{1}{N}\sum_{\beta=1}^{N}U_\beta^4$$

对于 $i\neq j$

$$E(u_i^3u_j)=\frac{-1}{N(N-1)}\sum_{\beta=1}^{N}U_\beta^4$$

$$E(u_i^2u_j^2)=\frac{1}{N(N-1)}\Big(\sum_{\beta=1}^{N}U_\beta^2\Big)^2-\frac{1}{N(N-1)}\sum_{\beta=1}^{N}U_\beta^4$$

对于 $i\neq j\neq k$

$$E(u_i^2 u_j u_k)=\frac{2}{N(N-1)(N-2)}\sum_{\beta=1}^{N}U_\beta^4$$

$$-\frac{1}{N(N-1)(N-2)}\left(\sum_{\beta=1}^{N}U_\beta^2\right)^2$$

对于 $i\neq j\neq k\neq l$

$$E(u_i u_j u_k u_l)=\frac{-6}{N(N-1)(N-2)(N-3)}\sum_{\beta=1}^{N}U_\beta^4$$

$$+\frac{3}{N(N-1)(N-2)(N-3)}\left(\sum_{\beta=1}^{N}U_\beta^2\right)^2$$

从而

$$E(\bar{u}^4)=\frac{(N-n)[N^2-(6n-1)N+6n^2]}{n^3N(N-1)(N-2)(N-3)}\sum_{\beta=1}^{N}U_\beta^4$$

$$+\frac{3(n-1)(N-n)(N-n-1)}{n^3N(N-1)(N-2)(N-3)}\left(\sum_{\beta=1}^{N}U_\beta^2\right)^2$$

定理2.4 设 $u_1,\cdots,u_n$ 是来自总体 $U_1,\cdots,U_N$ 的随机样本，且 $\bar{U}=0$，则

$$E(\bar{u}^2)=O\left(\frac{1}{n}\right),\quad E(\bar{u}^4)=O\left(\frac{1}{n^2}\right)$$

证明 由定理2.3(i)

$$E(\bar{u}^2)=\frac{N-n}{n(N-1)}\frac{1}{N}\sum_{\beta=1}^{N}U_\beta^2\leqslant\frac{1}{n}\sigma^2=O\left(\frac{1}{n}\right)$$

此处 $\sigma^2=\frac{1}{N}\sum_{\beta=1}^{N}(U_\beta-\bar{U})^2=\frac{1}{N}\sum_{\beta=1}^{N}U_\beta^2$.

由定理2.3 (iii)

$$E(\bar{u}^4)=\frac{(N-n)[N^2-(6n-1)N+6n^2]}{n^3N(N-1)(N-2)(N-3)}\sum_{\beta=1}^{N}U_\beta^4$$

$$+\frac{3(n-1)(N-n)(N-n-1)}{n^3N(N-1)(N-2)(N-3)}\left(\sum_{\beta=1}^{N}U_\beta^2\right)^2$$

记

$$\mu_4=\frac{1}{N}\sum_{\beta=1}^{N}U_\beta^4,\qquad \mu_2=\frac{1}{N}\sum_{\beta=1}^{N}U_\beta^2$$

则

$$E(\bar{u}^4)=\frac{1}{n^3}\frac{(N-n)[N^2-(6n-1)N+6n^2]}{(N-1)(N-2)(N-3)}\mu_4$$

$$+\frac{3}{n^2}\frac{(n-1)}{n}\frac{N}{(N-1)}\frac{(N-n)(N-n-1)}{(N-2)(N-3)}\mu_2^2$$

$$\leqslant\frac{1}{n^3}\mu_4+\frac{3}{n^2}\mu_2^2=O\left(\frac{1}{n^2}\right)$$

定理2.5 设随机变量序列[ξ_n]满足

$$E\xi_n=O\left(\frac{1}{n}\right),\qquad E\xi_n^2=O\left(\frac{1}{n}\right)$$

又设 $w_1,\cdots,w_n$ 是来自总体 $W_1,\cdots,W_N$ 的随机样本．则

$$E\left(\xi_n\sum_{i=1}^{n}w_i\right)=O(1)$$

证明

$$E\left(\xi_n\sum_{i=1}^{n}w_i\right)=E\left[\xi_n\sum_{i=1}^{n}(w_i-\bar{W})\right]+n\bar{W}E(\xi_n)$$

由假设知

$$n\bar{W}E(\xi_n)=O(1)$$

今令 $u_i=w_i-\bar{W}$，则

$$\left|E\left[\xi_n\sum_{i=1}^{n}(w_i-\bar{W})\right]\right|=\left|E\left(\xi_n\sum_{i=1}^{n}u_i\right)\right|=|nE(\xi_n\bar{u})|$$

$$\leqslant n\sqrt{E(\xi_n^2)}\sqrt{E(\bar{u}^2)}$$

$$=n\sqrt{O\left(\frac{1}{n}\right)}\sqrt{O\left(\frac{1}{n}\right)}=O(1)$$

§2.2 简单估值法(简记为SE法)

(一) 无偏估计量

定理2.6 设 $y_1,\cdots,y_n$ 是来自总体 $Y_1,\cdots,Y_N$ 的随机样本。记号 $\bar{Y}$, $\hat{Y}$, σ^2, S^2, C^2 同第一章§1.1. 又记

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i,\quad \hat{y}=N\bar{y}$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

则

(A_1) $E\bar{y}=\bar{Y}$ $(E\hat{y}=\hat{Y})$;

(A_2) $\sigma_{\bar{y}}^2=\frac{1-f}{n}S^2$ $\left(C_{\bar{y}}^2=\frac{1-f}{n}\frac{S^2}{\bar{Y}^2}=\frac{1-f}{n}C^2\right)$, 其中

$\sigma_{\bar{y}}^2=E(\bar{y}-\bar{Y})^2$, $f=\frac{n}{N}$, $C_{\bar{y}}^2=\frac{E(\bar{y}-\bar{Y})^2}{\bar{Y}^2}$,

(A_3) $E(s^2)=S^2$, $(E(s_{\bar{y}}^2)=\sigma_{\bar{y}}^2)$, 其中 $s_{\bar{y}}^2=\frac{1-f}{n}s^2$.

证明 证(A_1):

设第 i 个入样号码为 θ_i, 即 $y_i=Y_{\theta_i}$. 于是

$$E(\bar{y})=E\left(\frac{1}{n}\sum_{i=1}^{n}Y_{\theta_i}\right)=\frac{1}{n}\sum_{i=1}^{n}E(Y_{\theta_i})$$

$$=\frac{1}{n}\sum_{i=1}^{n}\sum_{\beta=1}^{N}Y_\beta P\{\theta_i=\beta\}=\frac{1}{n}\sum_{i=1}^{n}\sum_{\beta=1}^{N}Y_\beta\frac{1}{N}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\bar{Y}=\bar{Y}$$

证(A_2): 令

$$U_i=Y_i-\bar{Y},\ i=1,\cdots,N;\quad u_i=y_i-\bar{Y},\ i=1,\cdots,n$$

显然, $\bar{U}=0$, $u_1,\cdots,u_n$ 是来自总体 $U_1,\cdots,U_N$ 的随机样本, $\bar{u}=\bar{y}-\bar{Y}$. 根据定理2.3(i), 有

$$\sigma_{\bar{y}}^2 = E(\bar{y}-\bar{Y})^2 = E(\bar{u}^2) = \frac{N-n}{nN(N-1)}\sum_{\beta=1}^{N} U_\beta^2$$

$$= \frac{N-n}{nN}\,\frac{1}{N-1}\sum_{\beta=1}^{N}(Y_\beta-\bar{Y})^2 = \frac{1-f}{n}S^2$$

证(A_3):

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(u_i-\bar{u})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}u_i^2 - \frac{n}{n-1}\bar{u}^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} u_i^2 - \bar{u}^2\right]$$

根据定理2.2及2.3，有

$$E(s^2) = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}E(u_i^2) - E(\bar{u}^2)\right]$$

$$= \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}E(U_{\beta_i}^2) - \frac{1-f}{n}S^2\right]$$

$$= \frac{n}{n-1}\left[\frac{1}{N}\sum_{\beta=1}^{N}U_\beta^2 - \frac{1-f}{n}S^2\right]$$

$$= \frac{n}{n-1}\left[\frac{N-1}{N}S^2 - \frac{N-n}{nN}S^2\right] = S^2$$

定理2.6告诉我们，用

$\bar{y}$ 估 $\bar{Y}$ （$\hat{\bar{y}}$ 估 $\bar{Y}$）

s^2 估 S^2

$s_{\bar{y}}^2$ 估 $\sigma_{\bar{y}}^2$（估计 $\bar{y}$ 的方差）

这些估计是无偏的.

从定理2.6(A_2)看到无放回抽样以 $\bar{y}$ 估 $\bar{Y}$ 时，其均方误差为

$$\sigma_{\bar{y}}^2 = E(\bar{y}-\bar{Y})^2 = \frac{1-f}{n}S^2$$

而有放回抽样以 $\bar{y}$ 估 $\bar{Y}$ 时，其均方误差为

$$E(\bar{y}-\bar{Y})=E\left(\frac{1}{n}\sum_{i=1}^{n} y_i-\bar{Y}\right)^2=\frac{1}{n^2}E\left[\sum_{i=1}^{n}(y_i-\bar{Y})\right]^2$$

$$=\frac{1}{n^2}E\left[\sum_{i=1}^{n}(y_i-\bar{Y})^2+\sum_{i\neq j}(y_i-\bar{Y})(y_j-\bar{Y})\right]$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}E(y_i-\bar{Y})^2=\frac{1}{n}\,\frac{N-1}{N}S^2=\frac{1-1/N}{n}S^2$$

所以，无放回抽样优于有放回抽样.

（二）区间估计

当 n 相当大时，$\bar{y}$ 接近 $N(\bar{Y},\ \sigma^2_{\bar{y}})$ 分布律①．故给定 $e>0$，

---

① 其理论根据如下：

定理 设有限总体序列为

$$\pi_N(Y_{N1}, Y_{N2}, \cdots, Y_{NN}),\quad N=1,2,\cdots$$

令

$$\bar{Y}_N=\frac{1}{N}\sum_{i=1}^{N}Y_{Ni},\qquad S_N^2=\frac{1}{N-1}\sum_{i=1}^{N}(Y_{Ni}-\bar{Y}_N)^2$$

且对 $r=3$，4 及 N 充分大时有

$$\frac{\frac{1}{N}\sum_{i=1}^{N}(Y_{Ni}-\bar{Y}_N)^r}{\left[\frac{1}{N}\sum_{i=1}^{N}(Y_{Ni}-\bar{Y})^2\right]^{r/2}}=O(1)$$

又设 $(y_{N1},\cdots,y_{Nn})$ 是来自 π_N 的随机样本，其中 n 的大小依赖 N，使得 $\lim\limits_{N\to\infty}\frac{N}{n}=c$（有限常数），则

$$\lim_{N\to\infty}P\left[\frac{\bar{y}_N-\bar{Y}_N}{\sqrt{\sigma^2_{\bar{y}_N}}}<x\right]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{1}{2}t^2}\,dt$$

其中 $\bar{y}_N=\frac{1}{n}\sum_{i=1}^{n}y_{Ni}$，$\sigma^2_{\bar{y}_N}=\frac{1-f}{N}S_N^2$，$f=\frac{n}{N}$.

证明 参考 A. Wald and J. Wolfowitz (1944): "Statistical Tests Based on Permutations of the Observation." *Ann. Math. Stat.*, Vol **15**, pp 358—372. 此处叙述的定理是其特例.

根据这条定理，将要调查的 π_N 看成定理中总体序列中的一个，从而得 $\bar{y}$ 近似有 $N(\bar{Y},\sigma^2_{\bar{y}})$ 分布律.

查 $N(0,1)$表，可得 t_ε 使

$$1-\varepsilon=P\left\{\left|\frac{\bar{y}-\bar{Y}}{\sigma_{\bar{y}}}\right|<t_\varepsilon\right\}$$

$$=P\{\bar{y}-t_\varepsilon\sigma_{\bar{y}}<\bar{Y}<\bar{y}+t_\varepsilon\sigma_{\bar{y}}\}$$

$$=P\{\hat{y}-t_\varepsilon N\sigma_{\bar{y}}<\hat{Y}<\hat{y}+t_\varepsilon N\sigma_{\bar{y}}\}$$

在实际工作中以 $s_{\bar{y}}=\sqrt{s_{\bar{y}}^2}$ 代替 $\sigma_{\bar{y}}$. 即给出 $\bar{Y}$ 与 $\hat{Y}$ 的区间估计.

（三）根据对准确度的要求，决定样本量 n 的方法

对准确度的要求有两种提法：

(i)要求 $\sigma_{\bar{y}}=\sqrt{\sigma_{\bar{y}}^2}\leqslant\delta$（要求绝对精度不小于事先指定的 δ），即要求

$$\sigma_{\bar{y}}^2=\left(\frac{1}{n}-\frac{1}{N}\right)S^2\leqslant\delta^2$$

解不等式，得 n 必须满足不等式

$$n\geqslant\frac{1}{\frac{\delta^2}{S^2}+\frac{1}{N}}=\frac{NS^2}{N\delta^2+S^2}$$

若欲使

$$P\{|\bar{y}-\bar{Y}|\leqslant d\}=1-\varepsilon \quad (\varepsilon,\ d\text{ 事先指定})$$

只需取 $\delta=d/t_\varepsilon$，则

$$n\geqslant\frac{NS^2}{N\left(\frac{d}{t_\varepsilon}\right)^2+S^2}$$

即可.

(ii)要求 $C_{\bar{y}}<\gamma$（要求相对精度高于事先指定的 γ），即

$$C_{\bar{y}}^2=\left(\frac{1}{n}-\frac{1}{N}\right)C^2\leqslant\gamma$$

解得

$$n\geqslant\frac{1}{\frac{\gamma^2}{C^2}+\frac{1}{N}}=\frac{NC^2}{N\gamma^2+C^2}$$

欲使

$$P\left\{\left|\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right|<h\right\}=1-\varepsilon \quad (\varepsilon,\ h \text{ 事先指定})$$

只需取 $\gamma=h/t_\varepsilon$，则

$$n\geqslant\frac{NC^2}{N\left(\frac{h}{t_\varepsilon}\right)^2+C^2}$$

即可.

实际工作中，可根据过去的资料，或先进行少量抽样，预估出 S^2 或 C^2，从而粗略确定样本额 n.

（四）部分估计

问题的提法：总体 π_N 中有一个部分"子体" $\pi_{N'}$，它的个数 N' 为未知，但 π_N 中的任一个体是否属于 $\pi_{N'}$ 有明确的判别法. 今要调查这个"子体" $\pi_{N'}$ 中个体的数量指标之和.

例如，前述在 100 平方公里的地块内，调查蝗蝻的数量时，往往要就该地域内的荒地和耕种地分别估出蝗蝻的数量，这时全部100000000个个体中是耕种地的一平方米的小块组成一"子体"，是荒地的另一些小块组成另一"子体".

这一问题的解决办法如下：设"子体"（用其数量指标表示）为 $\pi_{N'}=\{Z_1,\cdots,Z_{N'}\}$，今将 $\pi_N-\pi_{N'}$ 中的个体，即 π_N 中不属于 $\pi_{N'}$ 的个体的数量指标皆取作零，则

$$\pi_N=\{Y_1,\cdots,Y_N\}=\{Z_1,\cdots,Z_{N'},\underbrace{0,\cdots,0}_{N-N'\text{个}}\}$$

此时，$\tilde{Y}=\sum_{i=1}^{N}Y_i=\sum_{i=1}^{N'}Z_i=\tilde{Z}$ 恰好是要调查的对象. 于是问题化为本节(一)的情况. 将本节(一)的结论照搬过来，得下列结果：

设样本是

$$(y_1,\cdots,y_n)=(z_1,\cdots,z_{n'},\underbrace{0,\cdots,0}_{n-n'\text{个}})$$

记

$$P=\frac{N'}{N},\qquad Q=1-P$$

$$p=\frac{n'}{n},\qquad q=1-p$$

我们有

$$\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i=\frac{1}{N}\sum_{i=1}^{N'}Z_i=\frac{N'}{N}\,\frac{1}{N'}\sum_{i=1}^{N'}Z_i=P\bar{Z}$$

$$\sigma_Y^2=\frac{1}{N}\Big(\sum_{i=1}^{N}Y_i^2-N(\bar{Y})^2\Big)=\frac{1}{N}\Big(\sum_{i=1}^{N'}Z_i-NP^2\bar{Z}^2\Big)$$

$$=\frac{1}{N}\Big[\sum_{i=1}^{N'}(Z_i-\bar{Z})^2+N'\bar{Z}^2-NP^2\bar{Z}^2\Big]$$

$$=\frac{1}{N}(N'\sigma_Z^2+NPQ\bar{Z}^2)=P\sigma_Z^2+PQ\bar{Z}^2$$

$$S_Y^2=\frac{N}{N-1}\sigma_Y^2=\frac{N}{N-1}\Big(\frac{N'-1}{N}S_Z^2+PQ\bar{Z}^2\Big)$$

$$C_Y^2=\frac{N}{N-1}\Big(\frac{N'-1}{N}C_Z^2+PQ\Big)\frac{1}{P^2}$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{n}\sum_{i=1}^{n'}z_i=\frac{n'}{n}\,\frac{1}{n'}\sum_{i=1}^{n'}z_i=p\bar{z}$$

$$\sigma_{\bar{y}}^2=\frac{1-f}{n}S_Y^2=\frac{1-f}{n}\,\frac{N}{N-1}\Big(\frac{N'-1}{N}S_Z^2+PQ\bar{Z}^2\Big)$$

$$C_{\bar{y}}^2=\frac{1-f}{n}C_Y^2=\frac{1-f}{n}\,\frac{N}{N-1}\Big(\frac{N'-1}{N}C_Z^2+PQ\Big)\cdot\frac{1}{P^2}$$

$$s_{\bar{y}}^2=\frac{1-f}{n}s^2=\frac{1-f}{n}\,\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$=\frac{1-f}{n}\,\frac{1}{n-1}\Big(\sum_{i=1}^{n'}z_i^2-np^2\bar{z}^2\Big)$$

$$=\frac{1-f}{n}\frac{1}{n-1}\left[\sum_{i=1}^{n'}(z_i-\bar{z})^2+n'\bar{z}^2-np^2\bar{z}^2\right]$$

$$=\frac{1-f}{n}\frac{1}{n-1}\left[\sum_{i=1}^{n'}(z_i-\bar{z})^2+npq\bar{z}^2\right]$$

特别地，若调查对象就是 N'，那只要令 $Z_i=1$ 就行了．此时有

$$\bar{Y}=P,\quad \tilde{Y}=N',\quad \sigma_Y^2=PQ$$

$$S_Y^2=\frac{N}{N-1}PQ,\quad C_Y^2=\frac{N}{N-1}\frac{Q}{P}$$

$$\bar{y}=p,\quad \tilde{y}=Np,\quad \sigma_{\bar{y}}^2=\frac{1-f}{n}\frac{N}{N-1}PQ$$

$$s_{\bar{y}}^2=\frac{1-f}{n-1}pq,\quad C_{\bar{y}}^2=\frac{1-f}{n}\frac{N}{N-1}\frac{Q}{P}$$

因此，以

p	估	P
Np	估	N'
$\frac{N}{n}\sum_{i=1}^{n'}z_i(=Np\bar{z}=N\bar{y})$	估	$\tilde{Z}(=\tilde{Y})$

（五）考虑调查费用决定样本量之方法举例

例1 设调查费用为 c_0+c_1n．即每调查一个样需费用 c_1，而 c_0 为调查的基本费用．又设以 $\bar{y}$ 估 $\bar{Y}$，由于误差 $|\bar{y}-\bar{Y}|$ 所造成的损失为

$$\Lambda E(\bar{y}-\bar{Y})^2=\Lambda\left(\frac{1}{n}-\frac{1}{N}\right)S^2$$

则要使总耗费最小，即 $c_0+c_1n+\Lambda\left(\frac{1}{n}-\frac{1}{N}\right)S^2$ 最小，应取

$$n=\sqrt{\frac{\Lambda S^2}{c_1}}$$

例2 设调查费用为 c_0+c_1n，误差 $|\bar{y}-\bar{Y}|$ 所造成的损失为 $\Lambda E|\bar{y}-\bar{Y}|$．为使总耗费 $c_0+c_1n+\Lambda E|\bar{y}-\bar{Y}|$ 最小，求 n．

此时可认为 $\bar{y}-\bar{Y}$ 近似遵从 $N\left(0,\frac{\sigma^2}{n}\right)$分布．于是

$$E|\bar{y}-\bar{Y}|=\sqrt{\frac{2}{\pi}}\frac{\sigma}{\sqrt{n}}\int_0^{\infty}xe^{-\frac{x^2}{2}}dx=\sqrt{\frac{2}{\pi}}\cdot\frac{\sigma}{\sqrt{n}}$$

要使 $c_0+c_1n+A\sqrt{\frac{2}{\pi}}\cdot\frac{\sigma}{\sqrt{n}}$ 最小，应取 $n=\left(\frac{A^2\sigma^2}{2\pi c_1^2}\right)^{\frac{1}{3}}$．

实际工作中，可用由历史资料或少量抽样资料预估出的S^2或σ^2 来决定 n．

（六）估计好坏的标准

如果以某个统计量 w 估计未知参数 W，则

$E(w-W)=Ew-W$　称为估计量的偏量；

$E(w-W)^2=\text{Var}\,w-[Ew-W]^2$　称为均方偏差；

$\frac{E(w-W)^2}{W^2}$　称为相对均方偏差．

可以提出下面两个原则：

（i）如果样本额增大，偏量与均方偏差同时变小，而且偏量比均方偏差的平方根变小得更快，则估计是可用的．本节 $\bar{y}$ 的偏量为零，均方偏差为

$$E(\bar{y}-\bar{Y})^2=\frac{1-f}{n}S^2$$

上述条件能满足．

（ii）比较两种估计的好坏，以它们的均方偏差的大小为标准．均方偏差小者为佳．

§2.3　比估值法（简记为RE法）

（一）问题的提法

设总体中每个个体有两个指标 X，Y，于是有

$$\begin{pmatrix} X_1, & X_2, & \cdots, & X_N \\ Y_1, & Y_2, & \cdots, & Y_N \end{pmatrix}$$

按照§1.1的记号，有

$$\bar{X},\quad \bar{\bar{X}},\quad \sigma_X^2,\quad S_X^2,\quad C_X^2$$
$$\bar{Y},\quad \bar{\bar{Y}},\quad \sigma_Y^2,\quad S_Y^2,\quad C_Y^2$$

令

$$\rho=\frac{\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{S_X^2\cdot S_Y^2}}$$

称为相关系数.

我们的调查目标是：

$$R=\frac{\bar{Y}}{\bar{X}}=\frac{\bar{\bar{Y}}}{\bar{\bar{X}}}$$

有时 $\bar{X}$ 是已知的，这时，估 R 与估 $\bar{Y}$ 是同一问题、有时 $\bar{X}$ 是未知的.

例1 调查某城市1981年的人口(Y)，相对于1955年人口(X)的比率. 1955年的人口是已知的.

例2 调查一片庄稼病株所占的比例，个体（人工单位面积）的两个指标是：

X 为株数，　Y 为病株数

调查目标为　$R=\frac{\bar{\bar{Y}}}{\bar{\bar{X}}}\left(=\frac{\bar{Y}}{\bar{X}}\right)$. 这种情况 $\bar{X}$ 是未知的.

（二）比估值法

定理2.7 设总体 $\begin{pmatrix} X_1, & \cdots, & X_N \\ Y_1, & \cdots, & Y_N \end{pmatrix}$, $X_i>0$, $\bar{Y}>0$.

令　$R=\frac{\bar{Y}}{\bar{X}}$. 样本为

$$\begin{pmatrix} x_1, & \cdots, & x_n \\ y_1, & \cdots, & y_n \end{pmatrix}$$

记　$r=\bar{y}/\bar{x}$. 则有

(A_4)　$E(r)=R+O\left(\frac{1}{n}\right)$;

(A_5) $E(r-R)^2=\frac{1-f}{n}\frac{1}{\overline{X}^2}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2$

$$+O\left(\frac{1}{n^{3/2}}\right)=O\left(\frac{1}{n}\right),$$

(A_6) $E\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]$

$$=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{n}\right);$$

(A_7) $E\left[\frac{1}{\bar{x}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]$

$$=\frac{1}{\overline{X}^2}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{n^{1/2}}\right).$$

证明 证(A_4)：

用§2.1定理2.4，考虑

$$r-R=\frac{\bar{y}-R\bar{x}}{\bar{x}}=\frac{\bar{y}-R\bar{x}}{\bar{x}}\cdot\frac{\overline{X}-\bar{x}+\bar{x}}{\overline{X}}$$

$$=\frac{(\bar{y}-R\bar{x})(\overline{X}-\bar{x})}{\bar{x}\overline{X}}+\frac{\bar{y}-R\bar{x}}{\overline{X}}$$

$$=-\frac{(\bar{y}-R\bar{x})(\bar{x}-\overline{X})}{\bar{x}\overline{X}}+\frac{\bar{y}-R\bar{x}}{\overline{X}}$$

两边取期望，移项得

$$E(r)=R-E\left[\frac{(\bar{x}-\overline{X})(\bar{y}-R\bar{x})}{\bar{x}\overline{X}}\right]$$

记 $X^0=\min\{X_1,\cdots,X_N\}$，于是 $\bar{x}\geqslant X^0>0$，从而

$$|E(r)-R|\leqslant E\left|\frac{(\bar{x}-\overline{X})(\bar{y}-R\bar{x})}{\bar{x}\overline{X}}\right|$$

$$\leqslant\frac{1}{\overline{X}X^0}E\{|\bar{x}-\overline{X}||\bar{y}-R\bar{x}|\}$$

$$\leqslant \frac{1}{\bar{X}\bar{X}^0}\sqrt{E(\bar{x}-\bar{X})^2}\cdot\sqrt{E(\bar{y}-R\bar{x})^2}$$

记

$$X_i-\bar{X}=U_i, \quad x_i-\bar{X}=u_i$$
$$Y_i-RX_i=V_i, \quad y_i-Rx_i=v_i$$

则

$u_1,\cdots,u_n$ 为来自总体 $U_1,\cdots,U_N$ 的样本，且 $\bar{U}=0$

$v_1,\cdots,v_n$ 为来自总体 $V_1,\cdots,V_N$ 的样本，且 $\bar{V}=0$

根据定理2.4，有

$$E(\bar{x}-\bar{X})^2=E(\bar{u}^2)=O\left(\frac{1}{n}\right)$$

$$E(\bar{y}-R\bar{x})^2=E(\bar{v}^2)=O\left(\frac{1}{n}\right)$$

因而

$$|E(r)-R|\leqslant\frac{1}{\bar{X}\bar{X}^0}\cdot\sqrt{O\left(\frac{1}{n}\right)}\sqrt{O\left(\frac{1}{n}\right)}=O\left(\frac{1}{n}\right)$$

证(A_5)：因为

$$(r-R)^2=\frac{(\bar{y}-R\bar{x})^2}{\bar{x}^2}=\frac{(\bar{y}-R\bar{x})^2}{\bar{x}^2}\cdot\frac{\bar{X}^2-\bar{x}^2+\bar{x}^2}{\bar{X}^2}$$

$$=\frac{(\bar{y}-R\bar{x})^2}{\bar{X}^2}-\frac{\bar{x}^2-\bar{X}^2}{\bar{x}^2\bar{X}^2}(\bar{y}-R\bar{x})^2$$

所以

$$E(r-R)^2=\frac{1}{\bar{X}^2}E(\bar{y}-R\bar{x})^2-E\left[\frac{\bar{x}^2-\bar{X}^2}{\bar{x}^2\bar{X}^2}(\bar{y}-R\bar{x})^2\right]$$

而根据定理2.6，有

$$\frac{1}{\bar{X}^2}E(\bar{y}-R\bar{x})^2=\frac{1}{\bar{X}^2}E(\bar{v}^2)=\frac{1}{\bar{X}^2}\,\frac{1-f}{n}S_V^2$$

$$=\frac{1}{\bar{X}^2}\,\frac{1-f}{n}\,\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2$$

记 $X^*=\max\{X_1,\cdots,X_N\}$. 则根据定理2.4，有

$$E\left[\frac{\bar{x}^2-\bar{X}^2}{\bar{x}^2\bar{X}^2}(\bar{y}-R\bar{x})^2\right]\leqslant E\frac{|\bar{x}-\bar{X}|\cdot|\bar{x}+\bar{X}|(\bar{y}-R\bar{x})^2}{\bar{x}^2\bar{X}^2}$$

$$\leqslant\frac{X^*+\bar{X}}{(X^0)^2\bar{X}^2}E[|\bar{x}-\bar{X}|\cdot|\bar{y}-R\bar{x}|^2]$$

$$\leqslant\frac{X^*+\bar{X}}{(X^0)^2\bar{X}^2}\sqrt{E(\bar{x}-\bar{X})^2}\sqrt{E(\bar{y}-R\bar{x})^4}$$

$$=\frac{X^*+\bar{X}}{(X^0)^2\bar{X}^2}\sqrt{E(\bar{u}^2)E(\bar{v}^4)}$$

$$=\frac{X^*+\bar{X}}{(X^0)^2\bar{X}^2}\sqrt{O\left(\frac{1}{n}\right)O\left(\frac{1}{n^2}\right)}=O\left(\frac{1}{n^{3/2}}\right)$$

因而

$$E(r-R)^2=\frac{1}{\bar{X}^2}\frac{1-f}{n}\cdot\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{n^{3/2}}\right)=O\left(\frac{1}{n}\right)$$

证(A_6):

$$E\left\{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right\}$$

$$=E\left\{\frac{1}{n-1}\sum_{i=1}^{n}[y_i-Rx_i-(r-R)x_i]^2\right\}$$

$$=\frac{1}{n-1}E\left\{\sum_{i=1}^{n}[v_i-(r-R)x_i]^2\right\}$$

$$=E\left\{\frac{1}{n-1}\sum_{i=1}^{n}v_i^2\right\}-\frac{2}{n-1}E\left\{(r-R)\sum_{i=1}^{n}v_ix_i\right\}$$

$$+\frac{1}{n-1}E\left\{(r-R)^2\sum_{i=1}^{n}x_i^2\right\}$$

而根据定理2.6和2.4，有

$$E\left\{\frac{1}{n-1}\sum_{i=1}^{n}v_i^2\right\}=E\left\{\frac{1}{n-1}\sum_{i=1}^{n}(v_i-\bar{v})^2+\frac{n}{n-1}\bar{v}^2\right\}$$

$$=E\left\{\frac{1}{n-1}\sum_{i=1}^{n}(v_i-\bar{v})^2\right\}+\frac{n}{n-1}E(\bar{v}^2)$$

$$=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{n}\right)$$

又若令

$$r-R=\xi_n,\quad V_iX_i=W_i,\quad v_ix_i=w_i$$

则 $w_1,\cdots,w_n$ 为来自 $W_1,\cdots,W_N$ 之样本. 由本定理之(A_4)与(A_5), 有

$$E\xi_n=E(r-R)=O\left(\frac{1}{n}\right)$$

$$E\xi_n^2=E(r-R)^2=O\left(\frac{1}{n}\right)$$

故由定理2.5, 有

$$\frac{2}{n-1}E\left[(r-R)\sum_{i=1}^{n}v_ix_i\right]=\frac{2}{n-1}E\left[\xi_n\sum_{i=1}^{n}w_i\right]=O\left(\frac{1}{n}\right)$$

又

$$\frac{1}{n-1}E\left\{(r-R)^2\sum_{i=1}^{n}x_i^2\right\}\leqslant\frac{1}{n-1}n(X^*)^2E(r-R)^2=O\left(\frac{1}{n}\right)$$

从而

$$E\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{n}\right)$$

证(A_7):

$$E\left[\frac{1}{\bar{x}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]$$

$$=E\left[\frac{1}{\bar{X}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2-\frac{\bar{x}^2-\bar{X}^2}{\bar{x}^2\bar{X}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]$$

$$=\frac{1}{\bar{X}^2}E\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]$$

$$-E\left[\frac{\bar{x}^2-\bar{X}^2}{\bar{x}^2\bar{X}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]$$

若记 $Y^*=\max\{Y_1,\cdots,Y_N\}$，则

$$\sum_{i=1}^{n}(y_i-rx_i)^2\leqslant 2\left[\sum_{i=1}^{n}y_i^2+r^2\sum_{i=1}^{n}x_i^2\right]$$

$$\leqslant 2\left[n(Y^*)^2+\left(\frac{Y^*}{X^0}\right)^2 n(X^*)^2\right]$$

$$=2n\left[(Y^*)^2+\left(\frac{X^*Y^*}{X^0}\right)^2\right]$$

$$\left|\frac{\bar{x}^2-\bar{X}^2}{\bar{x}^2\bar{X}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right|$$

$$\leqslant\frac{|\bar{x}-\bar{X}||\bar{x}+\bar{X}|}{(X^0)^2\bar{X}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2$$

$$\leqslant\frac{2n}{n-1}\left[(Y^*)^2+\left(\frac{X^*Y^*}{X^0}\right)^2\right]\cdot\frac{X^*+\bar{X}}{(X^0)^2\bar{X}^2}|\bar{x}-\bar{X}|$$

则

$$E\left[\frac{\bar{x}^2-\bar{X}^2}{\bar{x}^2\bar{X}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]$$

$$\leqslant\frac{2n}{n-1}\left[(Y^*)^2+\left(\frac{X^*Y^*}{X^0}\right)^2\right]\frac{X^*+\bar{X}}{(X^0)^2\bar{X}^2}E|\bar{x}-\bar{X}|$$

$$\leqslant\frac{2n}{n-1}\left[(Y^*)^2+\left(\frac{X^*Y^*}{X^0}\right)^2\right]\frac{X^*+\bar{X}}{(X^0)^2\bar{X}^2}\sqrt{E(\bar{x}-\bar{X})^2}$$

$$=O(1)\cdot\sqrt{O\left(\frac{1}{n}\right)}=O\left(\frac{1}{\sqrt{n}}\right)$$

从而

$$E\left[\frac{1}{\bar{x}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2\right]$$

$$=\frac{1}{\bar{X}^2}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2+O\left(\frac{1}{\sqrt{n}}\right)$$

定理2.7说明，以 r 估 R 是可用的，它符合§2.2(六)中之原则，其均方偏差为

$$E(r-R)^2\doteq\frac{1-f}{n}\frac{1}{\bar{X}^2}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2$$

根据定理2.7(A_7)，近似地可以用

$$\frac{1-f}{n}\frac{1}{\bar{x}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2$$

估 $E(r-R)^2$.

如果$\bar{X}$已知，则我们可用 $r\bar{X}$ 估 $\bar{Y}$，此时均方偏差为

$$E(r\bar{X}-\bar{Y})^2\doteq\frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2$$

根据定理2.7(A_8)，近似地可以用

$$\frac{1-f}{n}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2$$

估 $E(r\bar{X}-\bar{Y})^2$.

上述的估值法就是比估值法.

为了便于与简单估值法比较，我们给这些量以适当的符号，并给出一些简化的表达式.

$$\bar{y}_{RE}=r\bar{X}$$

$$\begin{aligned}\sigma^2_{RE}&=\frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2\\&=\frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}[Y_i-\bar{Y}-R(X_i-\bar{X})]^2\\&=\frac{1-f}{n}[S_Y^2-2R\rho S_XS_Y+R^2S_X^2]\\&=\frac{1-f}{n}\bar{Y}^2[C_Y^2-2\rho C_XC_Y+C_X^2]\end{aligned}$$

$$\sigma_r^2=\frac{1-f}{n}\frac{1}{\bar{X}^2}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2=\frac{1}{\bar{X}^2}\sigma_{RE}^2$$

$$=\frac{1-f}{n}R^2[C_Y^2-2\rho C_X C_Y+C_X^2]$$

$$s_{RE}^2=\frac{1-f}{n}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2$$

$$=\frac{1-f}{n}\frac{1}{n-1}\left[\sum_{i=1}^{n}y_i^2-2r\sum_{i=1}^{n}x_i y_i+r^2\sum_{i=1}^{n}x_i^2\right]$$

$$s_r^2=\frac{1-f}{n}\frac{1}{\bar{x}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2$$

$$=\frac{1}{\bar{x}^2}s_{RE}^2$$

$$C_{RE}^2=\frac{\sigma_{RE}^2}{\bar{Y}^2}=\frac{1-f}{n}[C_Y^2-2\rho C_X C_Y+C_X^2]$$

$$C_r^2=\frac{\sigma_r^2}{R^2}=C_{RE}^2$$

（三）比估值法与简单估值法之比较

为便于区别，将简单估值法之记号均加上下标SE，改变如下：

$$\bar{y}_{SE}=\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$$

$$\sigma_{SE}^2=\sigma_{\bar{y}}^2=\frac{1-f}{n}S_Y^2$$

$$C_{SE}^2=C_{\bar{y}}^2=\frac{1-f}{n}C_Y^2$$

$$s_{SE}^2=s_{\bar{y}}^2=\frac{1-f}{n}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

为估 $\bar{Y}$，使用简单估值法时，有估计量 $\bar{y}_{SE}$，其相对均方偏差为 $C_{SE}^2=\frac{1-f}{n}C_Y^2$. 使用比估值法，当 $\bar{X}$ 已知时，有估计量 $\bar{y}_{RE}$，相对均方偏差为

$$C_{RE}^2=\frac{1-f}{n}[C_Y^2-2\rho C_X C_Y+C_X^2]$$

定义 我们称

$$e=\frac{\sigma_{SE}^2-\sigma_{RE}^2}{\sigma_{RE}^2}=\frac{C_{SE}^2-C_{RE}^2}{C_{RE}^2}=\frac{2\rho C_X C_Y-C_X^2}{C_Y^2-2\rho C_X C_Y+C_X^2}$$

为比估值法对简单估值法的效率增量.

定理2.8 比估值法优于简单估值法，即 $e>0$ 的充分必要条件为

$$\rho>\frac{C_X}{2C_Y}$$

证明 显然.

由定理2.8可见，使用辅助指标 X 时，X 的变异系数 C_Y^2 愈小愈好，而 X 与调查对象 Y 的相关系数 ρ 则愈大愈好.

(四) 部分估计之比估值法

定理2.7′ 设总体 $\begin{pmatrix} X_1,\cdots,X_N \\ Y_1,\cdots,Y_N \end{pmatrix}$，$X_i\geqslant 0$，$\bar{Y}>0$. 令 $R=\bar{Y}/\bar{X}$. 样本为

$$\begin{pmatrix} x_1,\cdots,x_n \\ y_1,\cdots,y_n \end{pmatrix}$$

记 $r=\bar{y}/\bar{x}$；n' 为 $x_1,\cdots,x_n$ 中不为零的数的个数. 若样本额 n 增大时，恒有 $\frac{n'}{n}\geqslant \varepsilon>0$，则定理2.7之结论 (A_4)，(A_5)，(A_6)，(A_7) 仍然成立.

证明 从定理2.7的证明可以看出，条件 $X_i>0$ 是为了保证 $\bar{x}$ 有大于 0 的下界. 在本定理的条件 $X_i\geqslant 0$ 与 $\frac{n'}{n}\geqslant \varepsilon>0$ 的条件

下，$\bar{z}$ 也是有大于0的下界的．因而证明除上述这点外，其它完全与定理2.7相同．

如果在§2.2(四)之问题提法中，增加假定：子体额 N' 是已知的，则可采用比估值法作部分估计．

对每个个体给予指标如下：

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix}=\begin{cases} \begin{pmatrix} Z_i \\ 1 \end{pmatrix}, & \text{如果该个体在子体内} \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & \text{如果该个体不在子体内} \end{cases}$$

则 $\bar{X}=N'$ 已知，$\bar{Z}=\bar{Y}$ 是调查对象．

设样本为

$$\left[\begin{pmatrix} y_1 \\ x_1 \end{pmatrix},\cdots,\begin{pmatrix} y_n \\ x_n \end{pmatrix}\right]=\left[\begin{pmatrix} z_1 \\ 1 \end{pmatrix},\cdots,\begin{pmatrix} z_{n'} \\ 1 \end{pmatrix},\underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix},\cdots,\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{n-n'\text{个}}\right]$$

将本节(二)照搬过来得：

以 $\bar{y}_{RE}=r\,\bar{X}=\dfrac{\frac{1}{n}\sum\limits_{i=1}^{n'}z_i}{\frac{n'}{n}}\cdot\dfrac{N'}{N}=\dfrac{N'}{N}\bar{z}$ 估 $\bar{Y}$；

以 $N\bar{y}_{RE}=N'\bar{z}$ 估 $\bar{Z}(=\bar{Y})$；

以

$$s^2_{RE}=\frac{1-f}{n}\,\frac{1}{n-1}\left[\sum_{i=1}^{n}y_i^2-2r\sum_{i=1}^{n}x_i\,y_i+r^2\sum_{i=1}^{n}x_i^2\right]$$

$$=\frac{1-f}{n}\,\frac{1}{n-1}\left[\sum_{i=1}^{n'}z_i^2-2\bar{z}\sum_{i=1}^{n'}z_i+(\bar{z})^2n'\right]$$

$$=\frac{1-f}{n}\,\frac{1}{n-1}\left[\sum_{i=1}^{n'}z_i^2-n'\,\bar{z}^2\right]$$

估

$$\sigma^2_{RE}=\frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-RX_i)^2$$

$$=\frac{1-f}{n}\frac{1}{N-1}\left[\sum_{i=1}^{N}Y_i^2-2R\sum_{i=1}^{N}X_iY_i+R^2\sum_{i=1}^{N}X_i^2\right]$$

$$=\frac{1-f}{n}\frac{1}{N-1}\left[\sum_{i=1}^{N'}Z_i^2-N'\bar{Z}^2\right]$$

$$=\frac{1-f}{n}\frac{N'-1}{N-1}S_Z^2$$

这时效率增量为

$$e=\frac{\sigma^2_{SE}-\sigma^2_{RE}}{\sigma^2_{RE}}$$

$$=\frac{\frac{1-f}{n}\frac{N-1}{N}\left[\frac{N'-1}{N}S_Z^2+PQ\bar{Z}^2\right]-\frac{1-f}{n}\frac{N'-1}{N-1}S_Z^2}{\frac{1-f}{n}\frac{N'-1}{N-1}S_Z^2}$$

$$=\frac{\frac{1-f}{n}\frac{N}{N-1}PQ\bar{Z}^2}{\frac{1-f}{n}\frac{N'-1}{N-1}S_Z^2}=\frac{NPQ\bar{Z}^2}{(N'-1)S_Z^2}$$

（五）部分之部分的成数问题

问题提法：设总体 π_N 中有子体 $\pi_{N'}$，$\pi_{N'}$ 中又有子子体 $\pi_{N''}$. 额数 N 已知，而 N' 及 N'' 均未知. 调查目标为$\frac{N''}{N'}$.

如果我们给 π_N 的个体赋予指标

$$\begin{pmatrix}Y_i\\X_i\end{pmatrix}=\begin{cases}\begin{pmatrix}1\\1\end{pmatrix} & \text{该个体在 } \pi_{N''} \text{ 中}\\[2ex] \begin{pmatrix}0\\1\end{pmatrix} & \text{该个体在 } \pi_{N'} \text{ 中而不在 } \pi_{N''} \text{ 中}\\[2ex] \begin{pmatrix}0\\0\end{pmatrix} & \text{该个体不在 } \pi_{N'} \text{ 中}\end{cases}$$

于是

$$R=\frac{\bar{Y}}{\bar{X}}=\frac{N''}{N'}$$

恰好是调查对象．问题化为$\bar{X}$是未知的比估值问题，将本节(二)之结论搬过来即可．

设样本为

$$\underbrace{\binom{1}{1},\cdots,\binom{1}{1}}_{n''\text{个}},\ \underbrace{\binom{0}{1},\cdots,\binom{0}{1}}_{n'-n''\text{个}},\ \underbrace{\binom{0}{0},\cdots,\binom{0}{0}}_{n-n'\text{个}}$$

则应以

$$r=\frac{\bar{y}}{\bar{x}}=\frac{n''}{n'}\quad \text{估}\quad R=\frac{N''}{N'}$$

且

$$s_r^2=\frac{1-f}{n}\frac{1}{\bar{x}^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i-rx_i)^2$$

$$=\frac{1-f}{n}\frac{1}{\left(\frac{n'}{n}\right)^2}\frac{1}{n-1}\left[n''\left(1-\frac{n''}{n'}\right)^2+(n'-n'')\left(-\frac{n''}{n'}\right)^2\right]$$

$$=\frac{n(1-f)}{(n')^2}\frac{1}{n-1}\frac{n''(n'-n'')}{n'}$$

$$=\frac{(1-f)nn''(n'-n'')}{(n-1)(n')^3}$$

第三章　分层抽样法

§3.1　简单估值法（记为 SSE 法）

（一）简单估值法

设总体 π_N 分为 K 层

$$\pi_N=\pi_{N_1}\cup\pi_{N_2}\cup\cdots\cup\pi_{N_\nu}\cup\cdots\cup\pi_{N_K}$$

其中

$$\pi_{N_\nu}=\{Y_{\nu 1},\cdots,Y_{\nu N_\nu}\}$$

$$\nu=1,2,\cdots,K;\quad N=N_1+N_2+\cdots+N_K$$

对每层定义参数

$$\bar{Y}_\nu=\frac{1}{N_\nu}\sum_{i=1}^{N_\nu}Y_{\nu i},\quad \tilde{Y}_\nu=\sum_{i=1}^{N_\nu}Y_{\nu i}$$

$$\sigma_\nu^2=\frac{1}{N_\nu}\sum_{i=1}^{N_\nu}(Y_{\nu i}-\bar{Y}_\nu)^2$$

$$S_\nu^2=\frac{1}{N_\nu-1}\sum_{i=1}^{N_\nu}(Y_{\nu i}-\bar{Y}_\nu)^2$$

$$C_\nu^2=S_\nu^2/\bar{Y}_\nu^2,\quad \nu=1,2,\cdots,K$$

各符号意义与第一章完全相同，只不过多了标志层号的足码 ν.

令

$$W_\nu=N_\nu/N$$

$$\bar{Y}=\sum_{\nu=1}^{K}W_\nu\bar{Y}_\nu=\frac{1}{N}\sum_{\nu=1}^{K}\tilde{Y}_\nu$$

$$\tilde{Y}=N\bar{Y}=\sum_{\nu=1}^{K}\tilde{Y}_\nu$$

调查对象是 $\bar{Y}$ 或 $\tilde{Y}$.

现在从各层独立地抽取一组随机样本，记第 ν 层的样本如下，

样本额　n_v；

样本　$y_{v1},\cdots,y_{vn_v}$；

统计量　$\bar{y}_v=\frac{1}{n_v}\sum_{i=1}^{n_v}y_{vi}$，　$\hat{y}=N_v\bar{y}_v$，

$$s_{\bar{y}_v}^2=\frac{1}{n_v-1}\sum_{i=1}^{n_v}(y_{vi}-\bar{y}_v)^2.$$

并记

$$\bar{y}_{SSE}=\sum_{v=1}^{K}W_v\bar{y}_v=\frac{1}{N}\sum_{v=1}^{K}\hat{y}_v$$

$$s_{SSE}^2=\sum_{v=1}^{K}W_v^2s_{\bar{y}_v}^2$$

$$\sigma_{SSE}^2=E(\bar{y}_{SSE}-\bar{Y})^2$$

定理3.1　在上述假定下，有

(A_1')　$E(\bar{y}_{SSE})=\bar{Y}$；

(A_2')　$\sigma_{SSE}^2=\sum_{v=1}^{K}W_v^2\frac{1-f_v}{n_v}S_v^2$，其中　$f_v=n_v/N_v$；

(A_3')　$E(s_{SSE}^2)=\sigma_{SSE}^2$.

证明　由§2.2(A_1)与(A_2)，有

$$E\bar{y}_v=\bar{Y}_v,\quad \sigma_{\bar{y}_v}^2=\frac{1-f_v}{n_v}S_v^2$$

再注意到$\bar{y}_1,\cdots,\bar{y}_K$相互独立，即可得($A_1'$)与($A_2'$)。又由§2.2($A_3$)知

$$E(s_{\bar{y}_v}^2)=\sigma_{\bar{y}_v}^2$$

从而直接可得(A_3')。

根据定理3.1，分层抽样简单估值法就是以

$\bar{y}_{SSE}$　估　$\bar{Y}$

s_{SSE}^2　估　$\bar{y}_{SSE}$之均方偏差σ_{SSE}^2

（二）按比例分配样本额的分层抽样法，以及此法与不分层

随机抽样法之比较

所谓按比例分配样本额是指取

$$n_\nu=\frac{N_\nu}{N}\cdot n=W_\nu n,\quad n=n_1+\cdots+n_K$$

以此代入定理 3.1 中，均方偏差σ^2_{SSE}的公式，得

$$\sigma^2_{SSE}(\text{按比例})=\sum_{\nu=1}^{K}W_\nu^2\frac{1-f_\nu}{n_\nu}S_\nu^2$$

$$=\sum_{\nu=1}^{K}W_\nu^2\frac{1-\frac{1}{N_\nu}\cdot\frac{N_\nu}{N}n}{\frac{N_\nu}{N}n}S_\nu^2$$

$$=\frac{1-f}{n}\sum_{\nu=1}^{K}W_\nu S_\nu^2$$

其中 $f=n/N$。

按第二章的 SE 估值法，可从整个总体 π_N 中，抽取额数为 n 的随机样本来估计 $\bar{Y}$。比较 SE 法与 SSE 法(按比例)之优劣，就是比较 σ^2_{SE} 与 σ^2_{SSE}（按比例）。

因为

$$\sigma^2_{SE}=\frac{1-f}{n}S^2$$

而

$$S^2=\frac{1}{N-1}\sum_{\beta=1}^{N}(Y_\beta-\bar{Y})=\frac{1}{N-1}\sum_{\nu=1}^{K}\sum_{i=1}^{N_\nu}(Y_{\nu i}-\bar{Y})^2$$

$$=\frac{1}{N-1}\sum_{\nu=1}^{K}\sum_{i=1}^{N_\nu}(Y_{\nu i}-\bar{Y}_\nu+\bar{Y}_\nu-\bar{Y})^2$$

$$=\frac{1}{N-1}\sum_{\nu=1}^{K}\sum_{i=1}^{N_\nu}(Y_{\nu i}-\bar{Y}_\nu)^2+\frac{1}{N-1}\sum_{\nu=1}^{K}N_\nu(\bar{Y}_\nu-\bar{Y})^2$$

$$=\frac{1}{N-1}\sum_{\nu=1}^{K}(N_\nu-1)S_\nu^2+\frac{N}{N-1}\sum_{\nu=1}^{K}W_\nu(\bar{Y}_\nu-\bar{Y})^2$$

其中

$$\frac{1}{N-1}\sum_{\nu=1}^{K}\sum_{i=1}^{N_\nu}(Y_{\nu i}-\bar{Y}_\nu)^2=\frac{1}{N-1}\sum_{\nu=1}^{K}(N_\nu-1)S_\nu^2$$

可看作层内误差项；

$$\frac{1}{N-1}\sum_{\nu=1}^{K}N_\nu(\bar{Y}_\nu-\bar{Y})^2=\frac{N}{N-1}\sum_{\nu=1}^{K}W_\nu(\bar{Y}_\nu-\bar{Y})^2$$

可看作层间误差项.

所以，SSE 法对 SE 法的效率增量为

$$e=\frac{\sigma_{SE}^2-\sigma_{SSE}^2}{\sigma_{SSE}^2}$$

$$=\frac{\frac{N}{N-1}\sum_{\nu=1}^{K}W_\nu(\bar{Y}_\nu-\bar{Y})^2-\sum_{\nu=1}^{K}\left(W_\nu-\frac{N_\nu-1}{N-1}\right)S_\nu^2}{\sum_{\nu=1}^{K}W_\nu S_\nu^2}$$

$$=\frac{\frac{N}{N-1}\sum_{\nu=1}^{K}W_\nu(\bar{Y}_\nu-\bar{Y})^2-\frac{1}{N-1}\sum_{\nu=1}^{K}(1-W_\nu)S_\nu^2}{\sum_{\nu=1}^{K}W_\nu S_\nu^2}$$

上式中，分子的第二项

$$\sum_{\nu=1}^{K}\left(W_\nu-\frac{N_\nu-1}{N-1}\right)S_\nu^2=\sum_{\nu=1}^{K}\left(\frac{N_\nu}{N}-\frac{N_\nu-1}{N-1}\right)S_\nu^2$$

是很小的，因此，效率增量为正数的可能性很大．一般说来，可见分层抽样法（按比例）要比不分层好．

如果将分子的第二项改为0，第一项之$\frac{N}{N-1}$改为1，则

$$e\doteq\frac{\sum_{\nu=1}^{K}W_\nu(\bar{Y}_\nu-\bar{Y})^2}{\sum_{\nu=1}^{K}W_\nu S_\nu}$$

由此式可见，层平均的差别愈大，即层间差异愈大，而各层内方差愈小，则效率增量愈大. 所以，若总体 π_N 中的个体间有明显差异，则应将性质相近的划归同一层，将 π_N 分成若干层，采用分层抽样.

（三）用SSE法（按比例分配样本额）时，决定最小样本总额问题

由(二)知

$$\sigma^2_{\hat{S}SE}=\frac{1-f}{n}\sum_{\nu=1}^{K}W_\nu S^2_\nu=\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{\nu=1}^{K}W_\nu S^2_\nu$$

记

$$C^2_{\hat{S}SE}=\frac{\sigma^2_{\hat{S}SE}}{\bar{Y}^2}$$

欲使

$$\sigma^2_{\hat{S}SE}\leqslant\delta,\quad \delta\text{ 为事先指定之正数}$$

即

$$\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{\nu=1}^{K}W_\nu S^2_\nu\leqslant\delta$$

只要

$$n\geqslant\frac{1}{\dfrac{\delta}{\sum_{\nu=1}^{K}W_\nu S^2_\nu}+\dfrac{1}{N}}$$

若欲使

$$C^2_{\hat{S}SE}\leqslant\gamma,\quad \gamma\text{ 为事先指定之正数}$$

即

$$\left(\frac{1}{n}-\frac{1}{N}\right)\sum_{\nu=1}^{K}W_\nu S^2_\nu\leqslant\gamma\bar{Y}^2$$

只要

$$n \geqslant \frac{1}{\dfrac{v\bar{Y}^2}{\sum_{\nu=1}^{K} W_\nu S_\nu^2} + \dfrac{1}{N}}$$

对于其中的未知参数 $\bar{Y}$ 及 S_ν^2，在实际工作中，可用其他办法作出预估.

（四）样本额之理想分配

前面讨论了将样本总额 n，按比例分配于各层. 此种分配法，有某种直观之合理性，但此法未考虑各层方差之不同. 按理说，方差很小之层，可以少抽，而方差大之层，宜多抽. 故进一步探讨更合理之分配法如下.

各种分配法之好坏，取决于估计量 $\bar{y}_{SSE}$ 的方差的大小. 于是，理想分配样本额问题可提为下述问题：

在制约 $n_1 + \cdots + n_K = n$ 的条件下，求适当的 $n_1, \cdots, n_K$，使得

$$\sigma^2_{\bar{S}SE} = \sum_{\nu=1}^{K} W_\nu^2 \frac{1-f_\nu}{n_\nu} S_\nu^2 = \sum_{\nu=1}^{K} \left(\frac{1}{n_\nu} - \frac{1}{N_\nu}\right) W_\nu^2 S_\nu^2$$

达最小. 这是一个制约极值问题.

定理3.2 理想样本分配额为

$$n_\nu = n \cdot \frac{W_\nu S_\nu}{\sum_{\nu=1}^{K} W_\nu S_\nu}, \quad \nu = 1, \cdots, K$$

证明 用拉氏乘子法解极值问题.

令

$$f(n_1, \cdots, n_K; \lambda) = \sigma^2_{\bar{S}SE} + \lambda\left(\sum_{\nu=1}^{K} n_\nu - n\right)$$

$$= \sum_{\nu=1}^{K} \left(\frac{1}{n_\nu} - \frac{1}{N_\nu}\right) W_\nu^2 S_\nu^2 + \lambda\left(\sum_{\nu=1}^{K} n_\nu - n\right)$$

得正规方程

$$\begin{cases} \dfrac{\partial f}{\partial n_\nu}=0, & \nu=1,\cdots,K \\ \dfrac{\partial f}{\partial \lambda}=0 \end{cases}$$

即

$$\begin{cases} \dfrac{W_\nu^2 S_\nu^2}{\lambda}=n_\nu^2, & \nu=1,\cdots,K \\ \sum\limits_{\nu=1}^{K} n_\nu-n=0 \end{cases}$$

解之即得

$$\begin{cases} \lambda=\left(\dfrac{1}{n}\sum\limits_{\nu=1}^{K} W_\nu S_\nu\right)^2 \\ n_\nu=\dfrac{W_\nu S_\nu}{\dfrac{1}{n}\sum\limits_{\nu=1}^{K} W_\nu S_\nu}=n\cdot\dfrac{W_\nu S_\nu}{\sum\limits_{\nu=1}^{K} W_\nu S_\nu} \end{cases}$$

系 当 S_ν 为常数时，

$$n_\nu=n\cdot\frac{W_\nu}{\sum\limits_{\nu=1}^{K} W_\nu}=n\cdot\frac{N_\nu}{N}$$

即为前述之按比例分配样本额．

要想真算出理想之样本分配额 n_ν，必须知道 S_ν(知道各 S_ν 之间的比例即可)．实际工作中 S_ν 的值是不知道的．但是有可能得到 S_ν 的近似比例，按近似的比例（即使是很粗糙的）来分配样本额，比臆断分配或按比例分配通常要高明得多．下面介绍几种获得 S_ν 的近似比例的方法．

(ⅰ)预查法

先在每一层中少量取一序样，各层分别为 m_1，…，m_K 个．设这些序样为

$$(z_{\nu 1},\cdots,z_{\nu m_\nu})\quad \nu=1,\cdots,K$$

以$\frac{1}{m_\nu-1}\sum_{i=1}^{m_\nu}(z_{\nu i}-\bar{z}_\nu)^2$近似代替$S_\nu^2$，求得理想分配样本额$n_1$，$\cdots$，$n_K$. 再抽样时，因为已经抽了序样，故只需分别在各层补抽$n_1-m_1,\cdots,n_K-m_K$个样就行了.

(ii)察往法

若总体过去曾被调查过，则现在按过去的S_ν的估计值分配样本额.

(iii)辅助指标法

总体π_N中的个体，除要调查的指标Y外，往往另有一个"规模指标X"，这种"规模指标"是度量个体的规模大小的. 例如，调查全国人口年出生率，调查个体为县，则各县的现有人口就是一个"规模指标". 假定分层是按"规模指标"的大小. 如调查人口年出生率时，将全国各县按拥有人口的多少分层. 在这种情况下，可以认为各层的变异系数大致相近. 即

$$S_\nu \propto \bar{Y}_\nu$$

(记号$\propto$表示S_ν与$\bar{Y}_\nu$成正比例)，亦即

$$N_\nu S_\nu \propto \tilde{Y}_\nu$$

又假定，指标X与Y差不多成正比例，即

$$\tilde{Y}_\nu \propto \tilde{X}_\nu$$

于是有

$$S_\nu \propto \bar{X}_\nu$$

从而可得理想分配样本额.

(五) 已知第ν层中每观测一个个体需费用F_ν. 当调查总费用固定为F时，求样本额之理想分配

此问题之解法与(四)类似.

此时制约条件变为$n_1F_1+\cdots+n_KF_K=F$，在此制约下，求$n_1,\cdots,n_K$使σ_{SSB}^2最小.

用拉氏乘子法解此制约极值问题，可得正规方程

$$\begin{cases} \dfrac{F_\nu W_\nu^2 S_\nu^2}{\lambda} = n_\nu^2 F_\nu^2, & \nu = 1, \cdots, K \\ \sum_{\nu=1}^{K} F_\nu n_\nu = F \end{cases}$$

解之可得

$$\lambda = \left(\frac{1}{F} \sum_{\nu=1}^{K} \sqrt{F_\nu} W_\nu S_\nu \right)^2$$

$$n_\nu = \frac{W_\nu S_\nu}{\sqrt{F_\nu} \frac{1}{F} \sum_{\nu=1}^{K} \sqrt{F_\nu} W_\nu S_\nu} = \frac{F}{F_\nu} \cdot \frac{\sqrt{F_\nu} W_\nu S_\nu}{\sum_{\nu=1}^{K} \sqrt{F_\nu} W_\nu S_\nu}$$

$$\nu = 1, \cdots, K$$

§3.2 比估值法（记为SRE法）

（一）比估值法

设总体的第 ν 层是

$$\begin{pmatrix} X_{\nu 1} \\ Y_{\nu 1} \end{pmatrix}, \cdots, \begin{pmatrix} X_{\nu N_\nu} \\ Y_{\nu N_\nu} \end{pmatrix}, \quad \nu = 1, \cdots, K$$

按照第二章的符号记法，各层相应的参数是

$$\begin{matrix} X_\nu, & \bar{X}_\nu, & \sigma_{X\nu}^2, & S_{X\nu}^2, & C_{X\nu}^2 \\ Y_\nu, & \bar{Y}_\nu, & \sigma_{Y\nu}^2, & S_{Y\nu}^2, & C_{Y\nu}^2 \\ \rho_\nu, & R_\nu & & & \end{matrix}$$

仅增加了标志层的足码.

另外还有总体的指标 X, $\bar{X}$; Y, $\bar{Y}$, 其意义同前. 调查对象仍然是 Y 或 $\bar{Y}$.

$$Y = \sum_{\nu=1}^{K} W_\nu Y_\nu$$

假定已知 $X_\nu(\nu = 1, \cdots, K)$, 则可以对每一$Y_\nu$ 用比估值法(RE法).

设第 ν 层的样本是

$$\begin{pmatrix} x_{\nu 1} \\ y_{\nu 1} \end{pmatrix}, \cdots, \begin{pmatrix} x_{\nu n_\nu} \\ y_{\nu n_\nu} \end{pmatrix}, \quad \nu = 1, \cdots, K$$

相应的统计量是 $\bar{x}_\nu$，$\bar{y}_\nu$，$r_\nu = \bar{y}_\nu / \bar{x}_\nu$.

假定每个层的样本额 n_ν 都足够大，以致于第二章§2.3的结论对每个层都适用，于是对第 ν 层用比估值法，以

$$r_\nu X_\nu \quad 估 \quad Y_\nu$$

其均方偏差近似为

$$\frac{1-f_\nu}{n_\nu}[S_{Y_\nu}^2 - 2R_\nu \rho_\nu S_{Y_\nu} S_{X_\nu} + R_\nu^2 S_{X_\nu}^2]$$

令

$$\bar{y}_{SRE} = \sum_{\nu=1}^{K} W_\nu r_\nu \bar{X}_\nu$$

$$s_{SRE}^2 = \sum_{\nu=1}^{K} W_\nu^2 \frac{1-f_\nu}{n_\nu} \frac{1}{n_\nu - 1}\Big(\sum_{i=1}^{n_\nu} y_{\nu i}^2 - 2r_\nu \sum_{i=1}^{n_\nu} x_{\nu i} y_{\nu i} + r_\nu^2 \sum_{i=1}^{n_\nu} x_{\nu i}^2\Big)$$

显然在上述假定下，有

$$E\bar{y}_{SRE} \doteq \bar{Y}$$

$\bar{y}_{SRE}$ 对 $\bar{Y}$ 的均方偏差 σ_{SRE}^2 近似为

$$\sum_{\nu=1}^{K} W_\nu^2 \frac{1-f_\nu}{n_\nu}[S_{Y_\nu}^2 - 2R_\nu \rho_\nu S_{Y_\nu} S_{X_\nu} + R_\nu^2 S_{X_\nu}^2]$$

而

$$E(s_{SRE}^2) \doteq \sum_{\nu=1}^{K} W_\nu^2 \frac{1-f_\nu}{n_\nu}[S_{Y_\nu}^2 - 2R_\nu \rho_\nu S_{Y_\nu} S_{X_\nu} + R_\nu^2 S_{X_\nu}^2]$$

因此，分层抽样比估值法(SRE)的法则是①，

---

① 若 $\bar{X}_\nu$ 未知，则以

$$\frac{\bar{y}_{SSE}}{\bar{x}_{SSE}} \quad 估 \quad \frac{\bar{Y}}{\bar{X}}$$

若 $\bar{X}_\nu$ 未知，而 $\bar{X}$ 已知，则以

$$\bar{X}\frac{\bar{x}_{SSE}}{\bar{y}_{SSE}} \quad 估 \quad \bar{Y}$$

以 $\bar{y}_{SRE}$ 估 $\bar{Y}$

以 s^2_{SRE} 估 σ^2_{SRE}

(二) SRE 法与 SSE 法之比较

因为 SSE 法就是对每一层用 SE 法，SRE 法就是对每一层用 RE 法. 所以我们可以写

$$\sigma^2_{SSE}=\sum_{\nu=1}^{K} W^2_\nu \sigma^2_{SE\nu}$$

$$\sigma^2_{SRE}=\sum_{\nu=1}^{K} W^2_\nu \sigma^2_{RE\nu}$$

如果对每个 ν，$\nu=1,\cdots,K$，都有 $\sigma^2_{RE\nu}\leqslant\sigma^2_{SE\nu}$，则

$$\sigma^2_{SRE}\leqslant\sigma^2_{SSE}$$

也就是说，若对每一层，比估值法都比简单估值法好，则对分层抽样来说，比估值法也比简单估值法好. 通常情况下，使用与 Y 关系密切的辅助指标 X，采用比估值法确实可以得到比较好的估计.

第四章　二阶抽样法

凡作调查必有一批完全明确之最小单位，它们组成总体．前几章所讨论之抽样法皆认为最小单位可直接编号，实际工作中往往遇到最小单位不能直接编号的情况．例如，要调查某省农户的某些指标，则各个农户即为最小单位．但全省农户数量太大，各农户的档案均归所在公社掌管，因而全省农户无法直接编号．此时，可将全省的公社编号作为第一性抽样单位，再在所抽的公社中将农户编号作为第二性抽样单位，采用二阶抽样法．二阶抽样的被抽个体集中于若干公社，采用随机抽样，被抽个体将遍布于全省各公社，因而二阶抽样较之随机抽样，在实际采样工作中，具有省人力、省时等优点．

§4.1　二阶抽样问题的一般提法

设总体 π_N 分为 K 组

$$\pi_N=\{\pi_{N_1},\cdots,\pi_{N_K}\}$$

对每个组 π_{N_i} $(i=1,\cdots,K)$ 都有一个指标 G_i (G_i 可以是 Y_i 或 $\bar{Y}_i\cdots$)，组内调查的目的是估计 G_i，而总的调查目的是要估计

$$G=G_1+\cdots+G_K$$

因为可能抽到任一组，从而可能在任一组内进行抽样调查，所以事先对每个组拟定一个抽样计划，并选定对组指标的估计量．

我们将上述有关的量，给以适当的符号，总结如下：

	$\overbrace{\pi_{N_1}, \quad \cdots, \quad \pi_{N_K}}^{\pi_N}$
组	π_{N_1}, $\cdots$, π_{N_K}
组 指 标	G_1, $\cdots$, G_K
拟定的组内抽样法	某法，$\cdots$，某法
拟 定 的 估 计 量	g_1, $\cdots$, g_K
估计量的均方偏差	$\sigma^2_{g_1}$, $\cdots$, $\sigma^2_{g_K}$
均方偏差的估计量	$s^2_{g_1}$, $\cdots$, $s^2_{g_K}$

我们假定所选之 g_i 是 G_i 的无偏估计量，$s^2_{g_i}$ 是 $\sigma^2_{g_i}$ 的无偏估计量．

此外，我们还需决定如何抽组．抽组法一经确定，即按抽组法抽组，然后对每个抽到的组按拟定的组内抽样法进行抽样，算出相应的估计量 g_i，最后根据二阶抽样所得的结果来估计 G，并研究其估计的误差．G 的估计依赖于抽组法，下一节我们将就下列两种抽组法讨论 G 的估计量．

（i）随机抽组法：等概无放回地抽组 k 次．

（ii）有返回$(p_1,\cdots,p_K)$抽组法：有放回地抽组 k 次，每次抽到组 π_{N_i} 的概率均为 p_i．

§4.2　二阶抽样之估值法则

（一）当抽组法为随机抽组法时之估值法

定理4.1　设用随机抽组法，抽到的 k 个组是

$$\pi_{N_{\theta_1}}, \cdots, \pi_{N_{\theta_k}}$$

则有

（A_1）　$g=\frac{K}{k}\sum_{i=1}^{k}g_{\theta_i}$ 是 G 的无偏估计量，即有

$$E(g)=G$$

（A_2）　g 的均方偏差为

$$\sigma_g^2=\frac{K^2}{k}V_1+\frac{K^2}{k}(1-f)V_2$$

其中

$$f=\frac{k}{K},\quad V_1=\frac{1}{K}\sum_{i=1}^{K}\sigma_{g_i}^2,\quad V_2=\frac{1}{K-1}\sum_{i=1}^{K}\left(G_i-\frac{G}{K}\right)^2$$

(A_{10}) $\hat{V}_1=\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}^2$ 是V_1的无偏估计量；

(A_{11}) $\hat{V}_2=\frac{1}{k-1}\sum_{i=1}^{k}\left(g_{\theta_i}-\frac{g}{K}\right)^2-\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}^2$ 是V_2的无偏估计量；

(A_{12}) 若$\sigma_{g\theta_i}^2$不存在无偏估计量$s_{g\theta_i}^2$，则以

$$\hat{\sigma}_g^2=\frac{K^2}{k}(1-f)\frac{1}{k-1}\sum_{i=1}^{k}\left(g_{\theta_i}-\frac{g}{K}\right)^2$$

来估计σ_g^2．这个估计是有偏的，但其偏与σ_g^2相比小于k/K．

证明 抽得之组号$\theta_1,\cdots,\theta_k$是随机变量．但对任何抽定之组号，$g_{\theta_1},\cdots,g_{\theta_k}$是相互独立的估计量．

我们用$E(t|\theta)$表示$\theta_1,\cdots,\theta_k$在取定的条件下t的数学期望，于是有

(i) $E(g_{\theta_i}|\theta)=G_{\theta_i},\ i=1,\cdots,k$；

(ii) $E(g_{\theta_i}^2|\theta)=\sigma_{g\theta_i}^2+G_{\theta_i}^2,\quad i=1,\cdots,k$；

(iii) $E(g_{\theta_i}g_{\theta_j}|\theta)=G_{\theta_i}G_{\theta_j},\quad i\neq j;i,j=1,\cdots,k.$

因此

$$E(g|\theta)=\frac{K}{k}(G_{\theta_1}+\cdots+G_{\theta_k})$$

$$E(g^2|\theta)=\frac{K^2}{k^2}\sum_{i=1}^{k}E(g_{\theta_i}^2|\theta)+\frac{K^2}{k^2}\sum_{i\neq j}E(g_{\theta_i}g_{\theta_j}|\theta)$$

$$=\frac{K^2}{k^2}\sum_{i=1}^{k}(\sigma_{g\theta_i}^2+G_{\theta_i}^2)+\frac{K^2}{k^2}\sum_{i\neq j}G_{\theta_i}G_{\theta_j}$$

$$=\frac{K^2}{k^2}\sum_{i=1}^{k}\sigma_{g\theta_i}^2+\frac{K^2}{k^2}\left(\sum_{i=1}^{k}G_{\theta_i}\right)^2$$

而g的无条件期望是

$$E(g)=E[E(g|\theta)]=E\left[\frac{K}{k}\sum_{i=1}^{k}G_{\theta_i}\right]$$

$$=\frac{K}{k}\sum_{i=1}^{k}EG_{\theta_i}$$

$$=\frac{K}{k}\sum_{i=1}^{k}\left[\sum_{j=1}^{K}G_jP(\theta_i=j)\right]$$

$$=\frac{K}{k}\sum_{i=1}^{k}\left[\sum_{j=1}^{K}G_j\cdot\frac{1}{K}\right]$$

$$=\sum_{j=1}^{K}G_j=G$$

这就是(A_1).

g^2的无条件期望是

$$E(g^2)=E[E(g^2|\theta)]$$

$$=\frac{K^2}{k^2}\sum_{i=1}^{k}E(\sigma^2_{g\theta_i})+\frac{K^2}{k^2}E\left(\sum_{i=1}^{k}G_{\theta_i}\right)^2$$

其中

$$E(\sigma^2_{g\theta_i})=\sum_{j=1}^{K}\sigma^2_{gj}P(\theta_i=j)=\frac{1}{K}\sum_{j=1}^{K}\sigma^2_{gj}=V$$

$$E\left(\sum_{i=1}^{k}G_{\theta_i}\right)^2=\sum_{i=1}^{k}E(G^2_{\theta_i})+\sum_{i\neq j}E(G_{\theta_i}G_{\theta_j})$$

$$=\sum_{i=1}^{k}\ \sum_{l=1}^{K}G_l^2P(\theta_i=l)$$

$$+\sum_{i\neq j}\ \sum_{l\neq m}G_lG_mP(\theta_i=l,\theta_j=m)$$

$$=\sum_{i=1}^{k}\ \sum_{l=1}^{K}G_l^2\frac{1}{K}+\sum_{i\neq j}\ \sum_{l\neq m}G_lG_m\frac{1}{K(K-1)}$$

$$=\frac{k}{K}\sum_{l=1}^{K} G_l^2+\frac{k(k-1)}{K(K-1)}\sum_{l\neq m}G_lG_m$$

$$=\frac{k}{K}\sum_{l=1}^{K} G_l^2+\frac{k(k-1)}{K(K-1)}\left[\left(\sum_{l=1}^{K}G_l\right)^2-\sum_{l=1}^{K}G_l^2\right]$$

$$=\left(\frac{k}{K}-\frac{k(k-1)}{K(K-1)}\right)\sum_{l=1}^{K} G_l^2+\frac{k(k-1)}{K(K-1)}\left(\sum_{l=1}^{K}G_l\right)^2$$

$$=\frac{k(K-k)}{K(K-1)}\sum_{l=1}^{K}G_l^2+\frac{k(k-1)}{K(K-1)}G^2$$

$$=\frac{k(K-k)}{K(K-1)}\left[\sum_{l=1}^{K} G_l^2-\frac{G^2}{K}\right]$$

$$+\left[\frac{k(K-k)}{K^2(K-1)}+\frac{k(k-1)}{K(K-1)}\right]G^2$$

$$=\frac{k(K-k)}{K(K-1)}\sum_{l=1}^{K}\left(G_l-\frac{G}{K}\right)^2$$

$$+\frac{k(K-k)+Kk(k-1)}{K^2(K-1)}G^2$$

$$=k\frac{1-f}{K-1}\sum_{l=1}^{K}\left(G_l-\frac{G}{K}\right)^2+\frac{k^2}{K^2}G^2$$

$$=k(1-f)V_2+\frac{k^2}{K^2}G^2$$

于是

$$E(g^2)=\frac{K^2}{k^2}\cdot kV_1+\frac{K^2}{k^2}\left[k(1-f)V_2+\frac{k^2}{K^2}G^2\right]$$

$$=\frac{K^2}{k}V_1+\frac{K^2}{k}(1-f)V_2+G^2$$

从而

$$\sigma_g^2=E(g^2)-(Eg)^2=\frac{K^2}{k}V_1+\frac{K^2}{k}(1-f)V_2$$

这就是 (A_9) 的结论.

现证 (A_{10}). s_{gj}^2 是 σ_{gj}^2 的无偏估计量，故

$$E(s_{g\theta_i}^2|\theta)=\sigma_{g\theta_i}^2$$

所以

$$E(\hat{V}_1)=E\left(\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}^2\right)=\frac{1}{k}\sum_{i=1}^{k}E(s_{g\theta_i}^2)$$

$$=\frac{1}{k}\sum_{i=1}^{k}E[E(s_{g\theta_i}^2|\theta)]$$

$$=\frac{1}{k}\sum_{i=1}^{k}E(\sigma_{g\theta_i}^2)$$

$$=\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{K}\sigma_{gj}^2P(\theta_i=j)$$

$$=\frac{1}{K}\sum_{j=1}^{K}\sigma_{gj}^2=V_1$$

证 (A_{11}),

$$E(\hat{V}_2)=E\frac{1}{k-1}\sum_{i=1}^{k}\left(g_{\theta_i}-\frac{g}{K}\right)^2-E\left(\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}^2\right)$$

$$=E\frac{1}{k-1}\left(\sum_{i=1}^{k}g_{\theta_i}^2-\frac{k}{K^2}g^2\right)-E\left(\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}^2\right)$$

$$=\frac{1}{k-1}\sum_{i=1}^{k}E(g_{\theta_i}^2)-\frac{k}{K^2(k-1)}E(g^2)$$

$$-E\left(\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}^2\right)$$

而

$$E(g_{\theta_i}^2)=E[E(g_{\theta_i}^2|\theta)]=E[\sigma_{g\theta_i}^2+G_{\theta_i}^2]$$
$$=E(\sigma_{g\theta_i}^2)+E(G_{\theta_i}^2)$$
$$=\frac{1}{K}\sum_{j=1}^{K}\sigma_{g_j}^2+\frac{1}{K}\sum_{j=1}^{K}G_j^2$$
$$=V_1+\frac{1}{K}\sum_{j=1}^{K}\left(G_j^2-\frac{G^2}{K}\right)+\frac{G^2}{K^2}$$
$$=V_1+\frac{K-1}{K}V_2+\frac{G^2}{K^2}$$

$$E(g^2)=\frac{K^2}{k}V_1+\frac{K^2}{k}(1-f)V_2+G^2$$

$$E\left(\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}\right)=V_1$$

所以有

$$E(\hat{V}_2)=\frac{k}{k-1}\left[V_1+\frac{K-1}{K}V_2+\frac{G^2}{K^2}\right]-\frac{k}{K^2(k-1)}\left[\frac{K^2}{k}V_1\right.$$
$$\left.+\frac{K^2}{k}(1-f)V_2+G^2\right]-V_1$$
$$=\left[\frac{k}{k-1}-\frac{1}{k-1}-1\right]V_1+\left[\frac{k}{k-1}\frac{K-1}{K}-\frac{1-f}{k-1}\right]V_2$$
$$=\left[\frac{k}{k-1}\frac{K-1}{K}-\frac{1-k/K}{k-1}\right]V_2$$
$$=\frac{k(K-1)-(K-k)}{K(k-1)}V_2=V_2$$

最后我们证(A_{12})：

$$E(\hat{\sigma}_g^2)=E\left[\frac{K^2}{k}(1-f)\frac{1}{k-1}\sum_{i=1}^{k}\left(g_{\theta_i}-\frac{g}{K}\right)^2\right]$$

(A_{15})　$\hat{\sigma}^2_{g'}=\dfrac{1}{k(k-1)}\sum\limits_{i=1}^{k}(Z_i-\bar{Z})^2$ 是 $\sigma^2_{g'}$ 的无偏估计量，

其中 $\bar{Z}=\dfrac{1}{k}\sum\limits_{i=1}^{k}Z_i$.

证明　因为抽组法是有返回$(p_1,\cdots,p_K)$，故 $\theta_1,\cdots,\theta_K$ 相互独立同分布 $P(\theta_i=j)=p_j$, $j=1,\cdots,K$；$i=1,\cdots,k$，从而$Z_1,\cdots,Z_k$ 相互独立同分布．于是有

$$E(g')=E\left(\frac{1}{k}\sum_{i=1}^{k}Z_i\right)=E(Z_1)$$

$$\sigma^2_{g'}=\mathrm{Var}(g')=\mathrm{Var}\left(\frac{1}{k}\sum_{i=1}^{k}Z_i\right)$$

$$=\left(\frac{1}{k}\right)^2\sum_{i=1}^{k}\mathrm{Var}(Z_i)=\frac{1}{k}\mathrm{Var}(Z_1)$$

由此知，欲证(A_{13})与(A_{14})只需计算$E(Z_1)$与 $\mathrm{Var}(Z_1)$

$$E(Z_1)=E\left(\frac{g_{\theta_1}}{p_{\theta_1}}\right)=E\left[E\left(\frac{g_{\theta_1}}{p_{\theta_1}}\middle|\theta\right)\right]$$

$$=E\left[\frac{1}{p_{\theta_1}}E(g_{\theta_1}|\theta)\right]=E\left[\frac{1}{p_{\theta_1}}G_{\theta_1}\right]$$

$$=\sum_{j=1}^{K}\frac{1}{p_j}G_jP(\theta_1=j)=\sum_{j=1}^{K}\frac{1}{p_j}G_j\cdot p_j$$

$$=\sum_{j=1}^{K}G_j=G$$

故有(A_{13})

$$E(g')=E(Z_1)=G$$

而

$$E(Z_1^2)=E\left[E\left\{\left(\frac{g_{\theta_1}}{p_{\theta_1}}\right)^2\middle|\theta\right\}\right]=E\left[\frac{1}{p^2_{\theta_1}}E(g^2_{\theta_1}|\theta)\right]$$

$$=E\left[\frac{1}{p_{\theta_1}^2}(\sigma_{\theta\theta_1}^2+G_{\theta_1}^2)\right]$$

$$=\sum_{j=1}^{K}\frac{1}{p_j^2}(\sigma_{\theta_j}^2+G_j^2)P(\theta_1=j)$$

$$=\sum_{j=1}^{k}\frac{1}{p_j}(\sigma_{\theta_j}^2+G_j^2)$$

$$\mathrm{Var}(Z_1)=E(Z_1^2)-(EZ_1)^2=\sum_{j=1}^{K}\frac{1}{p_j}(\sigma_{\theta_j}^2+G_j^2)-G^2$$

$$=\sum_{j=1}^{K}\frac{1}{p_j}\sigma_{\theta_j}^2+\sum_{j=1}^{K}p_j\left(\frac{G_j}{p_j}-G\right)^2$$

于是

$$\sigma_{\theta'}^2=\frac{1}{k}\mathrm{Var}(Z_1)$$

$$=\frac{1}{k}\sum_{j=1}^{k}\frac{1}{p_j}\sigma_{\theta_j}^2+\frac{1}{k}\sum_{j=1}^{k}p_j\left(\frac{G_j}{p_j}-G\right)^2$$

这就是(A_{14}).

(A_{15})是显然的，因为$\frac{1}{k-1}\sum_{i=1}^{K}(Z_i-\bar{Z})^2$是$\mathrm{Var}(Z_1)$的无偏估计量，所以

$$E(\hat{\sigma}_{\theta'}^2)=E\left(\frac{1}{k(k-1)}\sum_{i=1}^{k}(Z_i-\bar{Z})^2\right)$$

$$=\frac{1}{k}\mathrm{Var}(Z_1)=\sigma_{\theta'}^2$$

§4.3 组内为随机抽样时之估值法

前一节是二阶抽样的一般估值法则，不论指标G为何，不论组内抽样方案如何皆可用之．现设组内抽样为随机抽样，再分别

就两种抽组法列出相应的结论。

设

$$\pi_N=\pi_{N_1}\cup\cdots\cup\pi_{N_K}$$

$$\pi_{N_i}=(Y_{i1},\cdots,Y_{iN_i})\text{有样本}(y_{i1},\cdots,y_{in_i})$$

组指标
$$G_i=\tilde{Y}_i=\sum_{j=1}^{N_i}Y_{ij}$$

总的调查对象

$$G=\tilde{Y}=\tilde{Y}_1+\cdots+\tilde{Y}_K$$

G_i 的无偏估计量为

$$g_i=\tilde{y}_i=\frac{N_i}{n_i}\sum_{j=1}^{n_i}y_{ij}$$

g_i 的均方偏差为

$$\sigma_{g_i}^2=N_i^2\frac{1-f_i}{n_i}S_i^2$$

其中 $f_i=\frac{n_i}{N_i}$， $S_i^2=\frac{1}{N_i-1}\sum_{j=1}^{N_i}(Y_{ij}-\bar{Y}_i)^2$， $\bar{Y}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}Y_{ij}$.

$\sigma_{g_i}^2$ 的无偏估计量为

$$s_{g_i}^2=N_i^2\frac{1-f_i}{n_i}s_i^2$$

其中 $s_i^2=\frac{1}{n_i-1}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)$， $\bar{y}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}y_{ij}$.

（一）抽组法是随机抽组法（简记为 TSE）

由(A_8)知

$$\tilde{y}_{TSE}=\frac{K}{k}\sum_{i=1}^{k}\tilde{y}_{*i}$$

是 $\tilde{Y}=\tilde{Y}_1+\cdots+\tilde{Y}_K$ 的无偏估计量.

由(A_9)知

$$\sigma_{\tilde{y}_{TSE}}^2=\frac{K^2}{k}V_1+\frac{K^2}{k}(1-f)V_2$$

其中 $f=k/K$

$$V_1=\frac{1}{K}\sum_{i=1}^{K}\sigma_{\theta_i}^2=\frac{1}{K}\sum_{i=1}^{K}\frac{N_i^2}{n_i}(1-f_i)S_i^2$$

$$V_2=\frac{1}{K-1}\sum_{i=1}^{K}\left(G_i-\frac{G}{K}\right)^2=\frac{1}{K-1}\sum_{i=1}^{K}\left(\bar{Y}_i-\frac{\bar{Y}}{K}\right)^2$$

由(A_{10})知

$$\hat{V}_1=\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}^2=\frac{1}{k}\sum_{i=1}^{k}\frac{N_{\theta_i}^2}{n_{\theta_i}}(1-f_{\theta_i})s_{\theta_i}^2$$

是 V_1 的无偏估计量.

由(A_{11})知

$$\hat{V}_2=\frac{1}{k-1}\sum_{i=1}^{k}\left(g_{\theta_i}-\frac{g}{K}\right)^2-\frac{1}{k}\sum_{i=1}^{k}s_{g\theta_i}^2$$

$$=\frac{1}{k-1}\sum_{i=1}^{k}\left(\bar{y}_{\theta_i}-\frac{1}{k}\sum_{j=1}^{k}\bar{y}_{\theta_j}\right)^2$$

$$-\frac{1}{k}\sum_{i=1}^{k}\frac{N_{\theta_i}^2}{n_{\theta_i}}(1-f_{\theta_i})s_{\theta_i}^2$$

是 V_2 的无偏估计量.

若有

$$N_1=\cdots=N_K\equiv N_0$$

$$n_1=\cdots=n_K\equiv n_0$$

$$f_1=\cdots=f_K\equiv f_0=\frac{n_0}{N_0}$$

则上述估计量化简为

$$\bar{y}_{TSE}=\frac{KN_0}{k}\sum_{i=1}^{k}\bar{y}_{\theta_i}$$

$$\sigma_{\bar{y}_{TSE}}^2=\frac{K^2}{k}V_1+\frac{K^2}{k}(1-f)V_2$$

$$V_1=\frac{1}{K}\frac{N_0^2}{n_0}(1-f_0)\sum_{i=1}^{K}S_i^2$$

$$V_2=\frac{N_0^2}{K-1}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2$$

其中 $$\bar{Y}=\frac{1}{N}Y=\frac{1}{KN_0}Y$$

$$\hat{V}_1=\frac{N_0^2}{k}\sum_{i=1}^{k}s_{\bar{y}\theta_i}^2$$

$$\hat{V}_2=\frac{N_0^2}{k-1}\sum_{i=1}^{k}(\bar{y}_{\theta_i}-\bar{y})^2-\frac{N_0^2}{k}\sum_{i=1}^{k}s_{\bar{y}\theta_i}^2$$

其中

$$\bar{y}=\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}y_{\theta_i j}=\frac{1}{kN_0}\sum_{i=1}^{k}\bar{y}_{\theta_i}$$

$$s_{\bar{y}\theta_i}^2=\frac{1-f_{\theta_i}}{n_{\theta_i}}\frac{1}{n_{\theta_i}-1}\sum_{j=1}^{n_{\theta_i}}(y_{\theta_i j}-\bar{y}_{\theta_i})^2$$

$$=\frac{1-f_0}{n_0}\frac{1}{n_0-1}\sum_{j=1}^{n_0}(y_{\theta_i j}-\bar{y}_{\theta_i})^2$$

在此情况下，若调查的目标是$\bar{Y}$，则估值法为

$\bar{y}_{TSE}=\bar{y}$ 是$\bar{Y}$的无偏估计量；

$$\sigma_{\bar{y}_{TSE}}^2=\frac{1-f_0}{kn_0}\frac{1}{K}\sum_{i=1}^{K}S_i^2+\frac{1-\frac{k}{K}}{k}\frac{1}{K-1}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2;$$

$\frac{1}{k}\sum_{i=1}^{k}s_{\bar{y}\theta_i}^2$ 是$\frac{1-f_0}{n_0}\frac{1}{K}\sum_{i=1}^{K}S_i^2$ 的无偏估计；

$\frac{1}{k-1}\sum_{i=1}^{k}(\bar{y}_{\theta_i}-\bar{y})^2-\frac{1}{k}\sum_{i=1}^{k}s_{\bar{y}\theta_i}^2$ 是$\frac{1}{K-1}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2$ 的无偏估计．

现引进下列记号

$$S_{内}^2=\frac{1}{K}\sum_{i=1}^{K} S_i^2=\frac{1}{K}\sum_{i=1}^{K}\left[\frac{1}{N_0-1}\sum_{j=1}^{N_0}(Y_{ij}-\bar{Y}_i)^2\right]$$

可视为平均组内均方偏差；

$$S_{外}^2=\frac{1}{K-1}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2$$

可视为组间均方偏差；

$$s_{内}^2=\frac{1}{k}\sum_{i=1}^{k} s_{\theta_i}^2=\frac{1}{k}\sum_{i=1}^{k}\left[\frac{1}{n_0-1}\sum_{j=1}^{n_0}(y_{\theta_i j}-\bar{y}_{\theta_i})^2\right]$$

可视为样本的平均组内均方偏差；

$$s_{外}^2=\frac{1}{k-1}\sum_{i=1}^{k}(\bar{y}_{\theta_i}-\bar{y})^2$$

可视为样本的组间均方偏差.

则

$$E(s_{内}^2)=E\left(\frac{1}{k}\sum_{i=1}^{k}s_{\theta_i}^2\right)=E\left(\frac{1}{k}\sum_{i=1}^{k}\frac{n_0}{1-f_0}s_{\bar{y}_{\theta_i}}^2\right)$$

$$=\frac{n_0}{1-f_0}E\left(\frac{1}{k}\sum_{i=1}^{k}s_{\bar{y}_{\theta_i}}^2\right)=\frac{1}{K}\sum_{i=1}^{K}S_i^2=S_{内}^2$$

$$E\left(s_{外}^2-\frac{1-f_0}{n_0}s_{内}^2\right)$$

$$=E\left(\frac{1}{k-1}\sum_{i=1}^{k}(\bar{y}_{\theta_i}-\bar{y})^2-\frac{1}{k}\sum_{i=1}^{k}s_{\bar{y}_{\theta_i}}^2\right)$$

$$=\frac{1}{K-1}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2=S_{外}^2$$

$\bar{y}_{TSE}$ 的均方偏差可表为

$$\sigma_{\bar{y}_{TSE}}^2=\frac{1-f_0}{kn_0}S_{内}^2+\frac{1-k/K}{k}S_{外}^2$$

倘若有 $k\ll K$，$n_0\ll N_0$，则

$$\sigma^2_{\bar{y}_{TSE}} \doteq \frac{1}{kn_0} S^2_{内} + \frac{1}{k} S^2_{外}$$

其估计量为

$$\frac{1}{kn_0} s^2_{内} + \frac{1}{k}\left(s^2_{外} - \frac{1}{n_0} s^2_{内} \right) = \frac{1}{k} s^2_{外}$$

（二）抽组法是有返回$(p_1, \cdots, p_K)$法（简记为TPE）

$$\bar{y}_{TPE} = \frac{1}{k}(Z_1 + \cdots + Z_k) = \frac{1}{k}\left(\frac{\bar{y}_{\theta_1}}{p_{\theta_1}} + \cdots + \frac{\bar{y}_{\theta_k}}{p_{\theta_k}} \right)$$

是 $\bar{Y}$ 的无偏估计，其均方偏差为

$$\sigma^2_{\bar{y}_{TPE}} = \frac{1}{k} \sum_{i=1}^{K} \frac{1}{p_i} \frac{N_i^2(1-f_i)}{n_i} S_i^2 + \frac{1}{k} \sum_{i=1}^{K} p_i \left(\frac{\bar{Y}_i}{p_i} - \bar{Y} \right)^2$$

而 $\frac{1}{k(k-1)} \sum_{i=1}^{k} (Z_i - Z)^2$ 是 $\sigma^2_{\bar{y}_{TPE}}$ 的无偏估计量.

当 $N_1 = \cdots = N_K \equiv N_0$，$n_1 = \cdots = n_k = n_0$，则

$$\bar{y}_{TPE} = \frac{Z_1 + \cdots + Z_k}{k}$$

是 $\bar{Y}$ 之无偏估计量，其中 $z_i = N_0 \bar{y}_{\theta_i} / p_{\theta_i}$. 此时其均方偏差为

$$\sigma^2_{\bar{y}_{TPE}} = \frac{N_0^2(1-f_0)}{kn_0} \sum_{i=1}^{K} \frac{S_i^2}{p_i} + \frac{N_0^2}{k} \sum_{i=1}^{K} p_i \left(\frac{\bar{Y}_i}{p_i} - K\bar{Y} \right)^2$$

$\sigma^2_{\bar{y}_{TPE}}$ 的无偏估计量为

$$\hat{\sigma}^2_{\bar{y}_{TPE}} = \frac{1}{k(k-1)} \sum_{i=1}^{k} (Z_i - \bar{Z})^2$$

如果在此情况下，问题是估 $\bar{Y}$，则

$$\bar{y}_{TPE} = \frac{1}{k}\left(\frac{\bar{y}_{\theta_1}}{Kp_{\theta_1}} + \cdots + \frac{\bar{y}_{\theta_k}}{Kp_{\theta_k}} \right)$$

是 $\bar{Y}$ 的无偏估计，其均方偏差为

$$\sigma^2_{\bar{y}_{TPE}} = \frac{(1-f_0)}{kn_0K^2} \sum_{i=1}^{K} \frac{S_i^2}{p_i} + \frac{1}{k} \sum_{i=1}^{K} p_i \left(\frac{\bar{Y}_i}{Kp_i} - \bar{Y} \right)^2$$

$$\hat{\sigma}^2_{\bar{y}_{TPE}}=\frac{1}{k(k-1)}\sum_{i=1}^{k}\left(\frac{\bar{y}_{\theta i}}{Kp_{\theta i}}-\frac{1}{kK}\sum_{i=1}^{k}\frac{\bar{y}_{\theta i}}{p_{\theta i}}\right)^2$$

（三）TSE 法与 SE 法之比较

当抽组法为随机抽组法，组内亦为随机抽样，且有 $N_1=\cdots=N_K\equiv N_0, n_1=\cdots=n_k\equiv n_0$ 时，以 $\bar{y}_{TSE}=\frac{1}{kn_0}\sum_{i=1}^{k}\sum_{j=1}^{n_0}y_{\theta ij}$ 估 $\bar{Y}$，其均方偏差为

$$\sigma^2_{\bar{y}_{TSE}}=\frac{1-f_0}{kn_0}S^2_{内}+\frac{1-\frac{k}{K}}{k}S^2_{外}$$

而如果对 π_N 直接取额数为 kn_0 的随机样本，按SE估值法，则以 $\bar{y}_{SE}=\frac{1}{kn_0}\sum_{i=1}^{kn_0}y_i$ 估 $\bar{Y}$，其均方偏差为

$$\sigma^2_{SE}=\frac{1-\frac{kn_0}{KN_0}}{kn_0}S^2=\frac{1-\frac{kn_0}{KN_0}}{kn_0}\left[\frac{1}{KN_0-1}\sum_{\beta=1}^{KN_0}(Y_\beta-\bar{Y})^2\right]$$

$$=\frac{1-\frac{kn_0}{KN_0}}{kn_0}\left[\frac{1}{KN_0-1}\sum_{i=1}^{K}\sum_{j=1}^{N_0}(Y_{ij}-\bar{Y})^2\right]$$

$$=\frac{1-\frac{kn_0}{KN_0}}{kn_0}\left[\frac{1}{KN_0-1}\sum_{i=1}^{K}\sum_{j=1}^{N_0}(Y_{ij}-\bar{Y}_i)^2+\frac{1}{KN_0-1}\sum_{i=1}^{K}N_0(\bar{Y}_i-\bar{Y})^2\right]$$

$$=\frac{1-\frac{kn_0}{KN_0}}{kn_0}\left[\frac{(N_0-1)K}{KN_0-1}S^2_{内}+\frac{N_0(K-1)}{KN_0-1}S^2_{外}\right]$$

因而相应的效率增量为

$$e=-\frac{1}{kn_0}\left\{\frac{\left[\frac{N_0-n_0}{N_0}-\frac{(KN_0-kn_0)(N_0-1)}{N_0(KN_0-1)}\right]S^2_{\text{内}}}{\frac{1-n_0/N_0}{kn_0}S^2_{\text{内}}+\frac{1-k/K}{k}S^2_{\text{外}}}\right.$$

$$\left.+\frac{\left[\frac{(K-k)n_0}{K}-\frac{(KN_0-kn_0)(K-1)}{K(KN_0-1)}\right]S^2_{\text{外}}}{\frac{1-n_0/N_0}{kn_0}S^2_{\text{内}}+\frac{1-k/K}{k}S^2_{\text{外}}}\right\}$$

$$=-\frac{n_0(k-1)(N_0-1)-(n_0-1)N_0(K-1)}{N_0(KN_0-1)}$$

$$\times\frac{[S^2_{\text{内}}-N_0S^2_{\text{外}}]}{\frac{N_0-n_0}{N_0}S^2_{\text{内}}+\frac{(K-k)n_0}{K}S^2_{\text{外}}}$$

$$=-\frac{n_0(k-1)(N_0-1)-(n_0-1)N_0(K-1)}{N_0(K-1)}$$

$$\times\frac{[S^2_{\text{内}}-S^2]}{\frac{N_0-n_0}{N_0}S^2_{\text{内}}+\frac{(K-k)n_0}{K}S^2_{\text{外}}}$$

当 $n_0\geqslant 2$，且 $k\leqslant K/2$ 时，

$$n_0(k-1)(N_0-1)-(n_0-1)N_0(K-1)<0$$

故若

$$S^2_{\text{内}}>S^2$$

则效率增量为正；若

$$S^2_{\text{内}}<S^2$$

则效率增量为负。

第五章　集团抽样法和系统抽样法

§5.1　集团抽样法

设总体 π_N 分为 K 组

$$\pi_N=\pi_{N_1}\cup\pi_{N_2}\cup\cdots\cup\pi_{N_K}$$

第 i 组 π_{N_i} 称为第 i 个集团．调查的目标是 $\bar{Y}$(或 Y)．对K个组采用某种抽组法抽组，而在抽到的组内采用普查．易见集团抽样可视为组内抽样法为普查之二阶抽样法．

当总体 π_N 中之个体缺少必要的档案材料，无法以个体为单位取样时，常需采用集团抽样法．例如，调查某市1978年后结婚之妇女生育的子女数．则全市所有1978年后结婚之妇女组成总体 π_N，个体为 1978 年后结婚之各位妇女．然而由于没有 1978 年后结婚之妇女的档案，无法按个体取样，新建立这样的档案无异于进行一次普查．因而按户籍派出所之管辖区划分成K个集团（即全市共有 K 个派出所），采用集团抽样法，对被抽到的集团，按户籍册进行普查．

（一）集团随机抽样法之简单估值法（简记为 CSE）

我们继续沿用第四章之符号．调查目标为 $G=\bar{Y}$．

当抽组法为随机抽样法时，由于组内抽样法为普查，因而

$$g_i=G_i=\sum_{j=1}^{N_i}Y_{ij}$$

g_i 估 G_i 之均方偏差为

$$\sigma^2_{g_i}=0$$

估计量为

$$s^2_{g_i}=0$$

由第四章之(A_8)，(A_9)，(A_{11})可得：以

$$\hat{y}_{CSE}=\frac{K}{k}\sum_{i=1}^{k}g_{\theta i}=\frac{K}{k}\sum_{i=1}^{k}\sum_{j=1}^{N_{\theta i}}Y_{\theta ij}\quad 估\quad \tilde{Y}$$

其均方偏差为

$$\sigma^2_{\hat{y}_{CSE}}=\frac{K^2}{k}\left(1-\frac{k}{K}\right)V_2$$

$$=\frac{K^2}{k}\left(1-\frac{k}{K}\right)\frac{1}{K-1}\sum_{i=1}^{K}\left(\sum_{j=1}^{N_i}Y_{ij}-\frac{\tilde{Y}}{K}\right)^2$$

$\sigma^2_{\hat{y}_{CSE}}$的估计量为

$$s^2_{\hat{y}_{CSE}}=\frac{K^2}{k}\left(1-\frac{k}{K}\right)\hat{V}_2$$

$$=\frac{K^2}{k}\left(1-\frac{k}{K}\right)\frac{1}{k-1}\sum_{i=1}^{n}\left(g_{\theta i}-\frac{g}{K}\right)^2$$

$$=\frac{K^2}{k}\left(1-\frac{k}{K}\right)\frac{1}{k-1}\sum_{i=1}^{k}\left(\sum_{j=1}^{N_i}Y_{\theta ij}-\frac{1}{k}\sum_{l=1}^{k}\sum_{j=1}^{N_{\theta l}}Y_{\theta lj}\right)^2$$

如果 N 已知，则可调查 $\bar{Y}$. 以

$$\bar{y}_{CSE}=\frac{1}{N}\hat{y}_{CSE}\quad 估\quad \bar{Y}$$

其均方偏差为

$$\sigma^2_{\bar{y}_{CSE}}=\frac{1}{N^2}\sigma^2_{\hat{y}_{CSE}}$$

以

$$s^2_{\bar{y}_{CSE}}=\frac{1}{N^2}s^2_{\hat{y}_{CSE}}\quad 估\quad \sigma^2_{\bar{y}_{CSE}}$$

特别地，当各集团中之个体数额相等时，即$N_1=\cdots=N_K=N_0$，$N=KN_0$，则

$$\bar{y}_{CSE}=\frac{1}{N}\hat{y}_{CSE}=\frac{1}{KN_0}\frac{K}{k}\sum_{i=1}^{k}\sum_{j=1}^{N_0}Y_{\theta ij}$$

$$=\frac{1}{k}\sum_{i=1}^{k}\frac{1}{N_0}\sum_{j=1}^{N_0}Y_{\theta_i j}=\frac{1}{k}\sum_{i=1}^{k}\bar{Y}_{\theta_i}$$

$$\sigma^2_{\bar{y}_{CSE}}=\frac{1}{N^2}\sigma^2_{\hat{y}_{CSE}}=\frac{1}{N^2}\frac{K^2}{k}\left(1-\frac{k}{K}\right)\frac{1}{K-1}$$

$$\times\sum_{i=1}^{K}\left(\sum_{j=1}^{N_0}Y_{ij}-\frac{\bar{Y}}{K}\right)^2$$

$$=\frac{1}{k}\left(1-\frac{k}{K}\right)\frac{1}{K-1}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2$$

$$s^2_{\bar{y}_{CSE}}=\frac{1}{N^2}s^2_{\hat{y}_{CSE}}$$

$$=\frac{1}{N^2}\frac{K^2}{k}\left(1-\frac{k}{K}\right)\frac{1}{k-1}\sum_{i=1}^{k}\left(\sum_{j=1}^{N_0}Y_{\theta_i j}\right.$$

$$\left.-\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{N_0}Y_{\theta_i j}\right)^2$$

$$=\frac{1}{k}\left(1-\frac{k}{K}\right)\frac{1}{k-1}\sum_{i=1}^{k}\left(\bar{Y}_{\theta_i}-\frac{1}{k}\sum_{i=1}^{k}\bar{Y}_{\theta_i}\right)^2$$

（二）集团有返回($p_1,\cdots,p_K$)法之估值法（简记为CPE）

当抽组法为有返回（p_1，…，p_K）法时，由第四章之(A_{13})，(A_{14})，(A_{15})得：以

$$\bar{y}_{CPE}=g=\frac{1}{k}\sum_{i=1}^{k}\frac{\bar{Y}_{\theta_i}}{p_{\theta_i}}\quad 估\quad \bar{Y}$$

其均方偏差为

$$\sigma^2_{\bar{y}_{CPE}}=\frac{1}{k}\sum_{i=1}^{K}p_i\left(\frac{\bar{Y}_i}{p_i}-\bar{Y}\right)^2$$

$\sigma^2_{\bar{y}_{CPE}}$的无偏估计为

$$s^2_{\bar{y}_{CPE}}=\frac{1}{k(k-1)}\sum_{i=1}^{k}\left(\frac{\bar{Y}_{\theta_i}}{p_{\theta_i}}-\frac{1}{k}\sum_{i=1}^{k}\frac{\bar{Y}_{\theta_i}}{p_{\theta_i}}\right)^2$$

由 $\sigma^2_{\hat{y}_{CPE}}$ 的表达式易见，如果选择 $p_i=\dfrac{\tilde{Y}_i}{\tilde{Y}}$，这样的概率分配是最好的，此时 $\sigma^2_{\hat{y}_{CPE}}=0$．当 $\tilde{Y}_i(i=1,\cdots,K)$ 都是正数时，这样的选择是有意义的．实际工作中 $\tilde{Y}_i$ 当然是不知道的，如果其为已知，则根本无需作抽样调查．然而有时有已往的资料可以利用，根据这些资料按上式确定 p_i，比之盲目指定一组 $(p_1,\cdots,p_K)$ 要高明得多．

〔三〕集团随机抽样法之比估值法（简记为 CRE）

若以一集团为一(大)个体，π_N 视为由 K 个(大)个体组成之总体．第 i 个(大)个体的指标为

$$U_i=\sum_{j=1}^{N_i}Y_{ij},\quad i=1,\cdots,K$$

调查目标为

$$\tilde{U}=\sum_{i=1}^{K}U_i=\sum_{i=1}^{K}\sum_{j=1}^{N_i}Y_{ij}=\tilde{Y}$$

对此采用第二章之 SE 法，则以

$$\hat{u}_{SE}=K\bar{u}_{SE}=\frac{K}{k}\sum_{i=1}^{k}U_{\theta_i}=\frac{K}{k}\sum_{i=1}^{k}\sum_{j=1}^{N_{\theta_i}}Y_{\theta_i j}=\hat{y}_{CSE}$$

估

$$\tilde{U}=\tilde{Y}$$

其均方偏差为

$$\begin{aligned}\sigma^2_{\hat{u}_{SE}}&=K^2\sigma^2_{\bar{u}_{SE}}\\&=K^2\frac{1-\dfrac{k}{K}}{k}\frac{1}{K-1}\sum_{i=1}^{K}(U_i-\bar{U})^2\\&=\frac{K^2}{k}\left(1-\frac{k}{K}\right)\frac{1}{K-1}\sum_{i=1}^{K}\left(\sum_{j=1}^{N_i}Y_{ij}-\frac{\tilde{Y}}{K}\right)^2\\&=\sigma^2_{\hat{y}_{CSE}}\end{aligned}$$

此估值法就是(二)中之 CSE 法．

若各集团所包含的个体数额$N_1,\cdots,N_K$已知，从而

$$N=N_1+\cdots+N_K$$

也已知．我们可对第i个(大)个体使用辅助指标$X_i=N_i, i=1,\cdots,K$，采用第二章之RE法估值，即称之为集团随机抽样法之比估值法(CRE)．此时，

$$\bar{X}=N_1+\cdots+N_K=N$$

$$R=\frac{\bar{U}}{\bar{X}}=\frac{\bar{Y}}{N}=\bar{Y}$$

当集团抽样数额k相当大时，由第二章(A_4)知，以

$$\bar{y}_{CRE}=r=\frac{\sum_{i=1}^{k}\sum_{j=1}^{N_{\theta_i}}Y_{\theta_i j}}{\sum_{i=1}^{k}N_{\theta_i}} \quad 估 \quad \bar{Y}=R=\frac{\bar{Y}}{N}$$

由第二章(A_5)知，其均方偏差为

$$\begin{aligned}\sigma^2_{\bar{y}_{CRE}}&=E(r-R)^2\\&\doteq\frac{1-k/K}{k}\frac{1}{\left(\frac{1}{K}\sum_{i=1}^{K}N_i\right)^2}\frac{1}{K-1}\\&\quad\times\sum_{i=1}^{K}\left(\sum_{j=1}^{N_i}Y_{ij}-\frac{\bar{Y}}{N}N_i\right)^2\\&=\frac{K^2}{N^2}\frac{1-k/K}{k}\frac{1}{K-1}\sum_{i=1}^{K}\left(\sum_{j=1}^{N_i}Y_{ij}-N_i\bar{Y}\right)^2\end{aligned}$$

且可以用

$$\frac{K^2}{N^2}\frac{1-k/K}{k}\frac{1}{k-1}\sum_{i=1}^{k}\left(\sum_{j=1}^{N_{\theta_i}}Y_{\theta_i j}-\frac{\sum_{i=1}^{k}\sum_{j=1}^{N_{\theta_i}}Y_{\theta_i j}}{\sum_{i=1}^{k}N_{\theta_i}}\cdot N_{\theta_i}\right)^2$$

估 $\sigma^2_{\bar{y}_{CRE}}$.

若调查目标为 $\bar{Y}$，则以

$$\hat{y}_{CRE}=N\bar{y}_{CRE}=\frac{N}{\sum_{i=1}^{k}N_{\theta_i}}\sum_{i=1}^{k}\sum_{j=1}^{N_{\theta_i}}Y_{\theta_i j}\quad 估\quad Y$$

其均方偏差近似为

$$\frac{K^2(1-k/K)}{k}\frac{1}{K-1}\sum_{i=1}^{K}\left(\sum_{j=1}^{N_i}Y_{ij}-N_i\bar{Y}\right)^2$$

§5.2　系统抽样法

系统抽样法是选一正整数 K，将 π_N 中个体依次排列为

$$\begin{array}{cccc}
1, & 2, & \cdots, & K, \\
K+1, & K+2, & \cdots, & 2K \\
2K+1, & 2K+2, & \cdots, & 3K \\
\cdots\cdots & & &
\end{array}$$

直至 N 为止

对号码 $1,2,\cdots,K$ 作随机抽样，若 i 入样，则 $K+i,2K+i,\cdots$ 皆入样．若我们将同一列之个体的全体看作一个集团，即 $1,K+1,2K+1,\cdots$ 为第一个集团；$2,K+2,2K+2,\cdots$ 为第二个集团；$\cdots$；$K,2K,3K,\cdots$ 为第 K 个集团．则系统抽样可视为集团抽样．

π_N 中个体的额数 N 不一定正好是 K 的整数倍，因而各列的个体数有可能相差一个，但当各列的个体数超过50时$\left(即\frac{N}{K}\geqslant 50\right)$，这一个之差是可以忽略的．在实际的大规模调查中这一点是极易满足的．以下我们恒假定 $N=KN_0$，并就只抽一个系统样本的情形作一些讨论．

系统抽样的最大优点是实际进行抽样时非常方便．如在野外调查蝗蝻数量，从 100000 平方米的地块中抽取 1000 个一平米的

小块调查．利用系统抽样法，则可按某一路线前进，每前进一百米调查一平方米，走到地头时，按与原前进路线相距一米之平行路线返回．如此往返走遍整个地块．这要比按随机抽定的号码，挑取 1000 个一平方米的小块作调查方便得多．不仅野外调查如此，即使是室内抽取卡片，从一叠100000张卡片中，按随机数表抽取1000张，也要比每隔 100 张抽取一张麻烦得多．因而系统抽样常为实际工作者所乐于采用．

（一）系统抽样的估值法（简记为 SYE）

如前所述，系统抽样法可视为集团抽样数为 1（$k=1$）的集团抽样，且 $N_1=\cdots=N_K=N_0$，故其估值法如下：以

$$\bar{y}_{SYE}=\frac{1}{N_0}\sum_{j=1}^{N_0}Y_{\theta j}\quad 估\quad \bar{Y}$$

其中 θ 为从$(1,\cdots,K)$抽样的入样号码．$\bar{y}_{SYE}$的均方偏差为

$$\sigma^2_{\bar{y}_{SYE}}=\left(1-\frac{1}{K}\right)\frac{1}{K-1}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2=\frac{K-1}{K}S^2_{外}$$

其中

$$S^2_{外}=\frac{1}{K-1}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2$$

由于有

$$\sum_{i=1}^{K}\sum_{j=1}^{N_0}(Y_{ij}-\bar{Y})^2=\sum_{i=1}^{K}\sum_{j=1}^{N_0}(Y_{ij}-\bar{Y}_i)^2+\sum_{i=1}^{K}N_0(\bar{Y}_i-\bar{Y})^2$$

$$(N-1)S^2=K(N_0-1)S^2_{内}+N_0(K-1)S^2_{外}$$

其中 $S^2_{内}=\frac{1}{K}\sum_{i=1}^{K}\left[\frac{1}{N_0-1}\sum_{j=1}^{N_0}(Y_{ij}-\bar{Y}_i)^2\right]$，可得

$$\sigma^2_{\bar{y}_{SYE}}=\frac{K-1}{K}S^2_{外}$$

$$=\frac{K-1}{K}\frac{1}{N_0(K-1)}[(N-1)S^2-K(N_0-1)S_{内}^2]$$

$$=\frac{N-1}{KN_0}S^2-\frac{K(N_0-1)}{KN_0}S_{内}^2$$

$$=\frac{N-1}{N}S^2-\frac{K(N_0-1)}{N}S_{内}^2$$

若从 π_N 直接抽取一额数为 N_0 之随机样本，则按 SE 法估值，有以

$$\bar{y}_{SE} \quad 估 \quad \bar{Y}$$

其均方偏差为

$$\sigma_{SE}^2=\frac{1-N_0/N}{N_0}S^2=\frac{1-1/K}{N_0}S^2$$

比较 $\sigma_{\bar{y}_{SYE}}^2$ 与 σ_{SE}^2 可得下列结论：

当 $S_{内}^2>S^2$ 时，系统抽样法优于随机抽样法；

当 $S_{内}^2<S^2$ 时，随机抽样法优于系统抽样法；

当 $S_{内}^2=S^2$ 时，两种抽样法 $\bar{Y}$ 的估计量精度一样.

系统抽样法估值的精度，与抽样时各列内个体间的均方差异的平均 $S_{内}^2$ 的大小有关. 因而 π_N 中各个体的排列次序（哪些个体排在同一列），影响着系统抽样法估值的精度. 下面就几种情况作一些讨论.

（二）个体的指标与其次序有线性关系

此时，有关系式 $Y_i=a+\beta i$，$i=1, \cdots, N$. 作变换

$$U_i=\frac{Y_i-a}{\beta}$$

则

$$U_i=i$$

$$\bar{U}=\frac{1}{N}\sum_{i=1}^{N}i=\frac{N+1}{2}$$

$$S_U^2=\frac{1}{N-1}\sum_{i=1}^{N}(U_i-\bar{U})^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)^2$$

$$=\frac{1}{N-1}\left[\sum_{i=1}^{N}i^2-N\left(\frac{N+1}{2}\right)^2\right]$$

$$=\frac{1}{N-1}\left[\frac{N(N+1)(2N+1)}{6}-\frac{N(N+1)^2}{4}\right]$$

$$=\frac{N(N+1)}{12}$$

若选定K，对$(U_1,\cdots,U_N)$作系统抽样，则将$U_1,\cdots,U_N$排列如下：

1,	2,	3,	…,	K
$K+1$,	$K+2$,	$K+3$,	…,	$2K$
$2K+1$,	$2K+2$,	$2K+3$,	…,	$3K$
……	……	……	……	……
$(N_0-1)K+1$,	$(N_0-1)K+2$,	$(N_0-1)K+3$,	…,	N_0K

故各列之平均值为

$$\bar{Y}_1=\frac{1+(K+1)+\cdots+[(N_0-1)K+1]}{N_0}$$

$$=\frac{K[1+2+\cdots+(N_0-1)]+N_0}{N_0}$$

$$=\frac{K(N_0-1)}{2}+1$$

$$\bar{Y}_2=\frac{2+(K+2)+\cdots+[(N_0-1)K+2]}{N_0}=\frac{K(N_0-1)}{2}+2$$

……

$$\bar{Y}_K=\frac{K(N_0-1)}{2}+K=\frac{K(N_0+1)}{2}$$

各列内之均方差为

$$S_1^2=\frac{1}{N_0-1}\sum_{j=1}^{N_0}\left[(j-1)K+1-\frac{K(N_0-1)}{2}-1\right]^2$$

$$=\frac{K^2}{N_0-1}\sum_{j=1}^{N_0}\left[(j-1)-\frac{(N_0-1)}{2}\right]^2$$

$$=\frac{K^2}{N_0-1}\left[\sum_{j=1}^{N_0}(j-1)^2-N_0\left(\frac{N_0-1}{2}\right)^2\right]$$

$$=\frac{K^2}{N_0-1}\left[\frac{(N_0-1)N_0(2N_0-1)}{6}-\frac{N_0(N_0-1)^2}{4}\right]$$

$$=\frac{K^2N_0(N_0+1)}{12}$$

$$S_2^2=\frac{1}{N_0-1}\sum_{j=1}^{N_0}\left[(j-1)K+2-\frac{K(N_0-1)}{2}-2\right]^2$$

$$=\frac{K^2}{N_0-1}\sum_{j=1}^{N_0}\left[(j-1)-\frac{(N_0-1)}{2}\right]^2$$

$$=S_1^2=\frac{K^2N_0(N_0+1)}{12}$$

$$S_3^2=\cdots=S_K^2=S_1^2=\frac{K^2N_0(N_0+1)}{12}$$

$$S_{\text{内}U}^2=\frac{1}{K}\sum_{i=1}^{K}S_i^2=S_1^2=\frac{K^2N_0(N_0+1)}{12}=\frac{N(N+K)}{12}$$

而

$$Y=\alpha+\beta U,\quad S_Y^2=\beta^2S_U^2,\quad S_{\text{内}Y}^2=\beta^2S_{\text{内}U}^2$$

所以

$$S_Y^2=\beta^2\frac{N(N+1)}{12}<S_{\text{内}Y}^2=\beta^2\frac{N(N+K)}{12}\quad(K>1)$$

故知这类总体系统抽样优于随机抽样.

（三）个体的指标与其次序有某种周期性关系

这类总体系统抽样估值的精度(用估计量的均方偏差度量),与K的选取有很大关系. 举例说明如下:

设π_N之个体的指标以l为周期，呈下述形式($N=Ml$)

$$Y_1=1,\cdots,Y_l=l,Y_{l+1}=1,\cdots,$$

$$Y_{2l}=l,\cdots,Y_{(M-1)l+1}=1,\cdots,\ Y_{Ml}=l$$

如果我们作系统抽样，选定的K为l（或l的倍数），则各列内之个体指标全都是同一个数值. 因而$S^2_{内}=0$. 这样的系统样本与直接从π_N中随机取一额数为1的样本效果一样,精度很低.

若选定$K=l-1$，则系统抽样排列为

$$\begin{array}{ccccc} 1, & 2, & 3, & \cdots, & l-1 \\ l, & 1, & 2, & \cdots, & l-2 \\ l-1, & l, & 1, & \cdots, & l-3 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 2, & 3, & 4, & \cdots, & l \end{array}$$

这时不论抽到哪一列，均有

$$\bar{y}_{SYE}=\frac{1+2+\cdots+l}{l}=\bar{Y}$$

在实际工作中，象这样精确地呈周期性的资料是没有的，但或多或少具有一定周期性的资料是很多的. 例如，我国绝大多数企业、事业单位均实行每周休息一天的工作制和月工资制，因而许多公用事业和社会经济现象均或多或少以七天或一月呈周期性. 倘若要调查某路公共汽车的日客流量、某商店的日售货量等等资料时，切忌每隔七天或一月调查一天. 而应选择K使系统抽样的每一列均包含有星期一、二、…、日，包含月初、月中、月末等各种日期.

（四）个体的次序随机排列

对于π_N中个体的某一种固定的排列次序,系统抽样估值的精

度可能优于随机抽样，也可能劣于随机抽样，无法预言．但对π_N中N个个体的所有$N!$种排列次序而言，系统抽样的平均精度等于随机抽样的精度．推演如下：

$$\sigma^2_{\bar{y}_{SYE}}=\frac{1}{K}\sum_{i=1}^{K}(\bar{Y}_i-\bar{Y})^2=\frac{1}{K}\sum_{i=1}\bar{Y}_i^2-\bar{Y}^2$$

对$N!$种排列的平均记为$\bar{\sigma}^2_{\bar{y}_{SYE}}$，有

$$\bar{\sigma}^2_{\bar{y}_{SYE}}=\frac{1}{N!}\sum^{N!}\left[\frac{1}{K}\sum_{i=1}^{K}\bar{Y}_i^2-\bar{Y}^2\right]$$

$$=\frac{1}{N!}\left[\frac{1}{K}\sum^{N!}\sum_{i=1}^{K}\left(\frac{1}{N_0}\sum_{j=1}^{N_0}Y_{ij}\right)^2-N!\bar{Y}^2\right]$$

$$=\frac{1}{N!}\frac{1}{KN_0^2}\sum_{i=1}^{K}\sum^{N!}\left[\sum_{j=1}^{N_0}Y_{ij}^2+\sum_{j\neq l}Y_{ij}Y_{il}\right]-\bar{Y}^2$$

$$=\frac{1}{N!\,KN_0^2}\sum_{i=1}^{K}\left[\sum_{j=1}^{N_0}\sum^{N!}Y_{ij}^2+\sum_{j\neq l}\sum^{N!}Y_{ij}Y_{il}\right]-\bar{Y}^2$$

$$=\frac{1}{N!\,KN_0^2}\sum_{i=1}^{K}\left[\sum_{j=1}^{N_0}(N-1)!\sum^{N}Y_{\alpha\beta}^2\right.$$

$$\left.+\sum_{j\neq l}(N-2)!\sum^{N(N-1)}Y_{\alpha\beta}Y_{\gamma\delta}\right]-\bar{Y}^2$$

$$=\frac{1}{N!\,KN_0^2}\sum_{i=1}^{K}\left[N_0(N-1)!\sum^{N}Y_{\alpha\beta}^2\right.$$

$$\left.+N_0(N_0-1)(N-2)!\sum^{N(N-1)}Y_{\alpha\beta}Y_{\gamma\delta}\right]-\bar{Y}^2$$

$$=\frac{1}{NN_0}\sum^{N}Y_{\alpha\beta}^2+\frac{N_0-1}{N(N-1)N_0}\sum^{N(N-1)}Y_{\alpha\beta}Y_{\gamma\delta}-\bar{Y}^2$$

$$=\left[\frac{1}{NN_0}-\frac{N_0-1}{N(N-1)N_0}\right]\sum^{N}Y_{\alpha\beta}^2$$

$$+\frac{N_0-1}{N(N-1)N_0}\left(\sum^{N}Y_{\alpha\beta}\right)^2-\bar{Y}^2$$

$$=\frac{N-N_0}{NN_0(N-1)}\sum^{N}Y^2_{\alpha\beta}-\left[1-\frac{N(N_0-1)}{(N-1)N_0}\right]\bar{Y}^2$$

$$=\frac{N-N_0}{NN_0(N-1)}\sum^{N}Y^2_{\alpha\beta}-\frac{N-N_0}{(N-1)N_0}\bar{Y}^2$$

$$=\frac{N-N_0}{NN_0}\frac{1}{N-1}\left[\sum^{N}Y^2_{\alpha\beta}-N\bar{Y}^2\right]$$

$$=\frac{1-N_0/N}{N_0}S^2=\sigma^2_{\bar{y}_B}$$

其中

$\sum^{N!}$表对π_N中N个个体的$N!$种排列求和；

$\sum^{N}$表对$(\alpha\beta)$取遍π_N中N个个体求和；

$\sum^{N(N-1)}$表对$(\alpha\beta)$，$(\gamma\delta)$为从π_N中N个个体取两个个体之一切可能排列求和。